

JOURNAL OF EDUCATIONAL RESEARCH

EDITORIAL BOARD

A. S. BARR, Editor
Professor of Education, University of Wisconsin, Madison, Wisconsin

WILLIAM A. BROWNELL, Dean, School of Education, University of California, Berkeley, California. Education Psychology, Learning, and Teaching.

WILLARD C. OLSON, Dean of College of Education, University of Michigan, Ann Arbor, Michigan. Child Development.

ROBERT A. DAVIS, Professor of Educational Research, George Peabody College for Teachers, Nashville, Tennessee. Evaluative Techniques.

WILSON B. THIEDE, Associate Professor of Education, University of Wisconsin, Madison, Wisconsin. Book Review Editor.

GUY D. McGRATH, Dean, College of Education, Arizona State College, Tempe, Arizona. Secondary Education.

ARTHUR E. TRAXLER, Educational Records Bureau, New York 32, New York. Educational Measurement.

VOLUME LIII
September 1959 - May 1960

Copyright 1959-1960 by
DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin



JOURNAL *of* EDUCATIONAL RESEARCH

VOLUME 53

SEPTEMBER, 1959

NUMBER 1

CONTENTS

An Investigation of Six Correlates of Student Withdrawal from High School—Joseph C. Bledsoe	3
Analysis of Observed Traits of Teachers Rated Superior in Demonstrating Creativeness in Teaching— Jesse A. Bond.....	7
An Opinionaire on Why College Students Choose to Teach— J. Marc Jantzen.....	13
The Field Trip as a Supplement to Teaching: An Experi- mental Study—Joseph S. Hillson, Alexander A. Wylie, Wolf P. Wolfensberger.....	19
A Study of Twenty Slow Learners—Sister Agnes Jerome	23
Some Perceptual Changes During Sensitivity Training— Kaj Lohmann, John H. Zenger, Irving R. Weschler.....	28
The Relationship of Authoritarianism to Rejection— Jack R. Frymier.....	33
A Comparison of Normal and Emotionally Ill Children on the Kahn Test of Symbol Arrangement— Howard H. Fink, Theodore C. Kahn	35
A Study of Predictive Validity of the Minnesota Teacher Attitude Inventory—Harry P. Day.....	37
Book Reviews	39

DEMBAR PUBLICATIONS, INC.

Madison 3, Wisconsin

AN INVESTIGATION OF SIX CORRELATES OF STUDENT WITHDRAWAL FROM HIGH SCHOOL

JOSEPH C. BLEDSOE
University of Georgia

THE AGE OF automation and atomic energy requires a better educated populace. Not only must the quality of education be improved but the proportions of our youth who attain higher levels of education must be increased. High school education is considered by many persons to be the minimum for effective participation in our democratic society. It is obviously too optimistic to expect everyone to attain this level since not everyone has the necessary mental aptitude. Far too many able students, however, withdraw from high school prior to graduation. In the nation as a whole, approximately one half of the young people who enroll for high school withdraw (1). Hollingshead (2) reported an intensive study of one midwestern community which showed that withdrawing from school is related to the class structure with those in the lower classes withdrawing in much higher proportions. The adolescents themselves gave economic need, peer isolation, and discrimination by teachers as important reasons for withdrawing. Other surveys have indicated that the three main factors behind school withdrawal are low intelligence with resulting low academic ability, profound dissatisfaction with the school program or with teachers, and economic reasons (3, 4).

Space does not permit a detailed listing of findings of numerous helpful studies of "drop-outs" or early school leavers. Many of these studies are listed in the references appended to this study. (See items 5 through 16.) In summary, these studies point up four needs: 1) early recognition of factors in any given community which encourage boys and girls to drop out of school; 2) early identification of symptoms of withdrawal; 3) taking preventive steps with prospective school leavers as soon as possible, and 4) treating each potential drop-out as an individual and not as a part of a group.

Purposes of the Study

This study reports findings of a detailed survey of 247 students who withdrew during the three school years from September 1953 through May 1956 from

the eighth, ninth, and tenth grades of a high school in a small Georgia city whose population was approximately 20,000. The city serves both as the seat of county government and as trade center for an agricultural area and as an industrial (textile) center. Six elementary schools and one high school serve the community, which appears to be fairly typical of many communities in north Georgia and the Piedmont area of the southeast.

Questions concerning relationships between student withdrawal from high school and six factors have been raised. These factors are: 1) sex, 2) size of class in first through eighth grades, 3) stability of elementary enrollment, 4) parent occupation, 5) level of parent education, and 6) level of reading comprehension. These questions are more explicitly stated in the discussion of the findings.

Procedures and Sources of Data

Students who voluntarily withdrew from the eighth, ninth, and tenth grades in the subject school during the three-year period from September, 1953 to May, 1956 were identified by a careful study of the quarterly personnel records of the school. Lists of students withdrawing included those persons known to be living in the community but did not include those known to have transferred to other schools; names of students were verified by teachers who checked with students in the home-room groups. Sources of information included the quarterly and permanent records in the high school principal's office, records in the counselor's office, and personal information from teachers, the visiting teacher, and in some instances the students themselves. Data concerning class size and elementary school or schools attended were obtained from school records in the six feeder elementary schools and in the superintendent's office. The reading comprehension test from which scores were obtained from the counselor's records was the California Reading Achievement Test (Advanced), administered to all students at the ninth grade level.

Findings

Question No. 1. Is there a relationship between sex of student and withdrawal from high school before graduation?

A total of 985 boys and 979 girls were enrolled in the grades from which 146 boys and 101 girls dropped out. Boys comprised 50.2 per cent of the total enrollment. Among the drop-outs, however, 59.1 per cent were boys, an excess of 8.9 in expected per cent.

Question No. 2. Are mean sizes of classes in which drop-outs are enrolled different from mean sizes of all classes for grades one through eight? (Eight related sub-questions.)

Table I indicates the mean number of pupils and standard deviations for classes in which the drop-outs were enrolled and for all classes. In grades one, two, and three, the mean number of students in classes in which drop-out students were enrolled was greater than the mean size of all classes. For the other five grades, the differences favored the classes in which students dropping out were enrolled; that is to say, classes in which drop-outs were found averaged fewer students.

Question No. 3. Do students enrolled in only one feeder elementary school drop out of high school in the same proportion as those enrolled in two or more feeder elementary schools?

Of those students who were enrolled continuously in the same elementary school, 9.2 per cent withdrew from high school, whereas 35.3 per cent of students who were enrolled in two or more schools dropped out. Thus, 3.84 times as many students who shifted enrollment dropped out as did the stable group.

Question No. 4. Is student withdrawal from school before graduation related to occupation of students' parents?

With the frequency of students actually dropping out of school expressed as a ratio of the frequency in the student population, the ratios for the various broad classifications are as follows: professional and managerial, .09; agricultural, .65; homemaking, .92; sales, .39; clerical, .06; skilled labor, 1.02; unskilled labor, 2.64; and miscellaneous (unemployed, retired, unknown), 3.49. A ratio of 1.00 indicates that frequency observed is equal to frequency expected. Thus, students whose parents are engaged in professional, managerial, agricultural, clerical and sales work drop out in less-than-expected proportions, while students whose parents are unskilled laborers, retired, unemployed, or whose occupation is unknown drop out in greater proportions than to be expected by chance.

Question No. 5. Is withdrawal from high school before graduation related to level of education of the students' parents?

With the frequency of students actually withdrawing expressed as a ratio of the frequency expected in terms of the proportion in the student population, the ratios for the various classifications of years of

parent education completed are as follows: parents completed one through four years of schooling, 1.08; parents completed five or six years, 3.15; parents completed seven years, 4.86; parents completed eight years, .92; parents completed one, two or three years of high school, .39; parents completed high school, .01; parents attended college one or more years, .00 (no withdrawal observed).

Question No. 6. How does the mean reading comprehension score of student drop-outs compare with mean reading comprehension score of those who remain in school?

The mean reading comprehension score for students who dropped out of the ninth and tenth grades was 7.9; mean reading comprehension score for remaining ninth graders was 8.9.

Discussion

Since this group of withdrawing students was taken from a single high school, it may be considered as a total population from that school or as a non-random sample of "drop-outs from high school." Since the group cannot be considered to be representative of any known population, the findings reported, strictly speaking, do not permit generalization beyond the group. In applying the findings of an incidental group such as this one to another group, the reasoning is sound only when there are real similarities and no crucial differences between the particular groups involved.

The findings suggest, none the less, that boys may be more likely to drop out of high school than girls. Many questions are suggested. Is the school program better adapted to the needs of girls than of boys? What are the forces within the community which cause boys to withdraw in such relatively greater proportions? If over-sized classes (for example, more than 35 pupils) must occur, would not the preferred alternative be to have larger classes in some other grade? Is it reasonable to suggest that more help is needed at the earlier stages to enable the pupil to acquire the fundamental tools of learning? Would not more sensitivity on the part of teachers to the needs of mobile students, better communications between schools, less petty faultfinding as students transfer from one school to another, be constructive measures? The factors of parent occupation and level of parent education are probably closely related to the student's level of aspiration. School administrators and teachers should do whatever they possibly can to encourage able persons from lower occupational and educational strata to further their education. Serious studies of the need and the implementation of an adult education program on a local basis appears appropriate. Limited ability to read with understanding is not only ego-deflating but is seriously handicapping in the academic arena. Implications for improved teaching of reading at all levels as well as increased efforts at diagnosis of reading difficulties and remedial teaching are evident.

TABLE I

MEAN NUMBER OF PUPILS AND STANDARD DEVIATIONS FOR CLASSES
IN WHICH DROP-OUTS WERE ENROLLED FOR ALL CLASSES
FOR GRADES ONE THROUGH EIGHT

Grade	Mean No. Pupils in Classes where "Drop- out" was enrolled	Mean No. Pupils in all Classes
One	40.6 ± 7.0	37.4 ± 7.5
Two	38.8 ± 7.1	37.2 ± 6.7
Three	42.7 ± 8.1	39.6 ± 7.4
Four	35.2 ± 6.4	36.1 ± 5.7
Five	32.1 ± 5.3	34.4 ± 4.9
Six	31.8 ± 3.9	32.6 ± 3.5
Seven	31.1 ± 5.6	33.6 ± 6.5
Eight	33.5 ± 4.2	33.6 ± 4.0

REFERENCES

1. Reavis, William C., "Excerpts from the Annual Meeting of Secondary School Principals: The Lack of School Holding Power," School Review, Vol. 60 (April, 1952), pp. 189-191.
2. Hollingshead, August B., Elmtown's Youth, New York: John Wiley and Sons, Inc., 1949.
3. Dillon, Harold J., Early School Leavers: Major Educational Problems (Pub. 401), New York, National Labor Committee, 1949.
4. Johnson, Elizabeth S. and Legg, C. E., "Why Young People Leave School," Bulletin of the National Association of Secondary School Principals, Vol. 32 (November 1948), pp. 14-24.
5. Cook, Edward S., Jr., "An Analysis of Factors Related to Withdrawal from High School Prior to Graduation," (unpublished Ed. D. dissertation, The University of Georgia, 1953).
6. Dresher, Richard S., "Factors in Voluntary Drop-outs in the Public Schools of Detroit, Michigan," (unpublished Ed. D. dissertation, Oregon State College, 1953).
7. Gragg, William L., "School Leavers: They Can Be Spotted in Junior High School," Clearing House, Vol. 25 (October 1950): 72-74.
8. Gragg, William L., "Findings in Ithaca's Continuous Survey of Drop-outs," Clearing House, Vol. 26 (March 1952): 413-414.
9. Holbeck, Elmer S., "Seven Ways to Help Prevent Drop-outs," The Nation's Schools, Vol. 45 (May 1950): 35-36.
10. Mack, A. R., "A Study of Drop-outs," Bulletin of the National Association of Secondary School Principals Vol. 38 (February 1954): 49-51.
11. Nancarrow, James E., "Reducing Drop-outs," Bulletin of the National Association of Secondary School Principals Vol. 34 (December 1950): 183-188.
12. Patterson, Walter C., "What are the Major Causes of Student Drop-out and What Should the Schools Do about the Present Condition?" Bulletin of the National Association of Secondary School Principals Vol. 39 (April 1955): 85-88.
13. Sando, Rudolph F., "How to Make and Utilize Follow-up Studies of School Leavers," Bulletin of the National Association of Secondary School Principals Vol. 36 (March 1952): 66-74.
14. Snapp, Daniel W., "Why They Drop-out: Eight Clues to Greater Holding Power," Clearing House Vol. 27 (April 1953): 492-494.
15. Thomas, Robert J., "An Empirical Study of the High School Drop-outs in Regard to Ten Possibly Related Factors," Journal of Educational Sociology Vol. 28 (September 1954): 11-18.
16. Woellner, R. C., "Early School Leavers," School Review Vol. 59 (December 1951): 511-515.

ANALYSIS OF OBSERVED TRAITS OF TEACHERS RATED SUPERIOR IN DEMONSTRATING CREATIVENESS IN TEACHING

JESSE A. BOND
University of California
Los Angeles, California

The present study involves an examination of one portion of a previous study which presented an analysis of ratings of student teachers.¹ In the previous study mean scores were listed showing ratings of 855 student teachers at University of California at Los Angeles on each of 32 traits pertaining to teaching. The student teachers consisted of 266 in grades one to six inclusive at the elementary level and 589 in grades seven through twelve in secondary schools. The study disclosed that the group scored consistently high in personal characteristics such as courtesy, cooperation, appearance and conscientiousness and relatively low in some qualities pertaining specifically to the practices and procedures of teaching. The lowest of these was achievement of discipline through responsible citizenship. Next to this was, "using creativeness in teaching" and third from the bottom, especially for secondary teachers, was "using speech as an effective instrument in teaching." These evidently were difficult tasks which challenged all teachers to the utmost. But while mean rating scores of the entire group were low, some teachers were rated superior in them. It was found that 102 of the 855 were rated superior in "discipline" and 266 superior in "effectiveness of speech." What kind of person was the superior disciplinarian and what other competencies did he possess which may have enabled him to be especially effective in classroom control? Were there qualities possessed by some teachers which made them especially able to use oral communication to marked advantage in teaching? Reports of studies pertaining to both of these two questions were presented in previous issues of the *Journal of Educational Research*.²

Purpose of the Present Study

The purpose of this study was to find out what kind of student teachers were markedly creative in teaching. Despite the fact that the 855 students, as a group, were rated next to the lowest of 32 traits in "creativity," it was discovered that 158 or 26.8 percent of the secondary group and 87 or 32.7 percent of the elementary group were classified su-

perior in this teaching characteristic. The purpose of this study then was to ascertain what characteristics these 245 teachers possessed which have contributed to their strength in "creativity."

Creativeness of the teacher as interpreted by the teacher training staff of directors, supervisors and supervising teachers at the University of California, Los Angeles was a trait or competency composed of several parts, the most important of which was the ability to interpret for the pupil the work at hand or present it to him in such way as to stimulate his curiosity and imagination and cause him to seek applications in light of his own background and interests. The teacher preparation staff took the position that some teachers are much more able than others to cause their pupils to see that which is sufficiently novel and engaging in subject content to cause them to respond to its challenges with the full strength that they possess. This quality, "creativity", was held to be the chief difference between the good average teacher and the brilliant leader who stimulates his students to rise to the maximum in their learning.

Method

In order to ascertain how the "creative teacher" was like and different from other teachers, the 245 persons who had been rated superior in this quality were compared with the entire unselected group of 855 of which 589 were secondary teachers and 266 were elementary. Comparisons were made on the basis of mean scores computed from ratings of all student teachers on 32 items of a rating scale. Each student had been rated on each item according to the following: 1. fail, 2. pass, 3. average, 4. strong, 5. superior. The rating scale had been formulated by the directors and supervisors of student teaching and had been used and interpreted extensively for a number of years before data for this study were collected. It was assumed, therefore, that all supervisors were in fairly close agreement as to its meaning and use.

Mean scores were calculated for each of four

* All footnotes will be found at the end of the article.

groups of student teachers. These are listed in Table I as follows: (a) secondary student teachers who were rated "5" (superior) in "creativity," (b) all secondary student teachers, (c) elementary student teachers who were rated "5" (superior) in "creativity" and (d) all elementary student teachers. In column (a) of Table I the mean scores are listed in rank order for "creative" secondary teachers from "5.0," superior creativity to 4.28 "attention to health conditions." The latter score was the lowest on the list of 32. In columns (b), (c) and (d) scores are presented according to agreement with the item as listed in column (a). It may be noted that mean scores in Table I are primarily in the "4" category. Two prominent reasons for this were, first, that high standards of selection had excluded from student-teaching the majority of weaker students and second, student teachers who were making poor progress usually dropped the course before completing it. In Table II the same four groups of student teachers as in Table I are listed but in each column only numerals are used showing the rank order of each mean score. Identical numbers for some of the items indicate identical mean scores for them.

Student teachers who ranked high in "creativity" will frequently be referred to as the "creative group" in subsequent portions of this manuscript.

Findings

Comparison of the "Creative Group" and Unselected Group in Secondary Schools.

A comparison of figures in columns a and b of Table I discloses that the mean scores for all of the 32 items are larger for the creative group than for the unselected group. The range of scores from the highest (5.00) to the lowest (4.28) for the creative group is .72. The range for the unselected group is .77, extending from a high of 4.68 to a low of 3.91. While the extent of range for the two groups is similar, it may be noted that the scores for the creative group are much higher on the scale. The highest mean score for the "creative" group is .32 of a point higher than the top score of the unselected group and there is a difference of .37 of a point in favor of the creative group on the lowest score of each. Furthermore, the average of mean scores is higher by .44 of a point in favor of the creative group. Student teachers attaining a total score of 4.60 or more on the rating scale received the final mark of "A" in student-teaching. It was found that 148 or over 93 percent of the secondary student teachers in the creative group received this mark while slightly less than 30 percent of the unselected group of teachers were similarly graded. All of this clearly shows that student teachers at the secondary level who were rated superior in "creativity" were judged to be superior teachers. A question arising immediately is: does creativity on the part of the teacher affect his general competency and is his general competency a reliable measure of his creativity? To the first part of this question there must

be an affirmative answer particularly in the light of the fact that supervisors stressed creativity as one of the focal points distinguishing the very strong teacher. It was also found that students who were rated low in this quality rarely received high final grades in teaching. Inasmuch as all scores are relatively high for the creative group it is difficult to ascertain to what extent general competency was a measure of creativity. Undoubtedly there was some "halo" effect involved, namely, an effective student teacher appeared strong at all points and was rated accordingly.

An examination of the rank order of traits in columns (a) and (b) of Table II reveals some of the major differences between the two groups of secondary teachers and sheds some light on the special purpose of this study, namely, to ascertain some of the traits which characterize the creative teacher. Aside from item number one, there are no significant differences between the creative group and the unselected group of secondary student teachers in the rank order of the first eight items. In item number nine, however, "initiative in achieving objectives" there is a difference of eight in rank which suggests that the naturally creative teacher may have been somewhat more independent and more vigorous in carrying out a plan of action as well as having better understanding of what it was he was trying to accomplish. Assuming this was the case it would be reflected in more confidence, sureness and initiation. This supposition was particularly borne out by the wide differences between the creative and the unselected groups in the next two items on the list, namely, "resourcefulness in teaching" and skill in planning. The creative group had identical mean scores in these two traits. "Resourcefulness" is ranked 14 places higher on the creative teacher list of traits than on the list of traits for the unselected group. "Planning" is eight places higher with creative teachers. There appears to be a close relationship between these two traits and creativity.

The rank order of two other traits on the list is sufficiently different to merit attention. These are "evaluation" and "attention to supplies and equipment." These appear much higher, relatively, on the list of the unselected group. These same two items were found in a previous study to be similarly higher on the list of an unselected group when compared with a selected group of secondary student teachers who were superior in using speech as a teaching instrument. It is highly suggestive that teachers lacking strength in planning and teaching creatively, take refuge in working with supplies and equipment and in testing pupils on what they presumably have learned on their own initiative.

Comparison of the "Creative Group" and Unselected Group in Elementary Schools.

Mean scores for creative elementary student teachers listed in column (c) of Table I show a range of .83 from a top score of 5.00 in "creative ability" to a low score of 4.17 in "achievement of discipline."

TABLE I
SUMMARIZED RATINGS OF STUDENT TEACHERS BY MEAN SCORES

		Ratings from 2 (low) to 5 (high)			
		Secondary Schools Grades 7-12 Inclusive		Elementary Schools Grades 1-6 Inclusive	
		a	b	c	d
		Secondary Student Teachers Creative Ability	All Secondary Student Teachers	Elementary Student Teachers Creative Ability	All Elementary Student Teachers
		158*	589*	87*	266*
1.	Creative ability	5.0	3.93	5.0	3.99
2.	Wholesome philosophy of life	4.89	4.58	4.67	4.47
3.	Cooperation with colleagues	4.86	4.63	4.69	4.54
4.	Sincere courtesy of manner	4.85	4.68	4.70	4.60
5.	Social adjustment	4.82	4.44	4.70	4.41
6.	Professional mindedness	4.82	4.48	4.68	4.41
7.	Seeks to grow intellectually	4.82	4.44	4.72	4.44
8.	Conscientiousness in meeting responsibilities	4.78	4.51	4.64	4.41
9.	Initiative in achieving objectives	4.77	4.16	4.55	4.10
10.	Emotional balance	4.75	4.44	4.70	4.43
11.	Resourcefulness in teaching	4.75	4.06	4.53	4.09
12.	Skill in planning	4.75	4.15	4.59	4.17
13.	Health and health habits	4.71	4.44	4.67	4.36
14.	Appearance	4.69	4.45	4.60	4.37
15.	Effective teacher-pupil relationship	4.68	4.21	4.48	4.12
16.	Understands content of his field as needed in schools	4.68	4.20	4.52	4.11
17.	Understands social values and implications of his field	4.68	4.26	4.56	4.23
18.	Constructive leadership	4.66	4.13	4.44	4.06
19.	Judicious selection and use of instructional materials	4.66	4.13	4.47	4.06
20.	Adaptability to a variety of teaching situations	4.65	4.11	4.61	4.16
21.	Sees relationships of his field to all fields	4.62	4.16	4.46	4.32
22.	Success in carrying out plans	4.61	4.10	4.50	4.12
23.	Understanding of young people	4.60	4.16	4.44	4.11
24.	Skill in evaluation of pupils' growth and achievement	4.57	4.18	4.39	4.08
25.	Forcefulness without offensive aggressiveness	4.52	3.98	4.42	4.11
26.	Possesses a broad cultural understanding	4.52	4.14	4.31	4.08
27.	Attention to supplies and equipment	4.51	4.18	4.27	4.09
28.	Provision for individual differences	4.48	4.03	4.41	4.12
29.	Attention to English skills as related to subject	4.43	3.94	4.40	4.17
30.	Effectiveness of speech	4.37	3.94	4.41	4.11
31.	Achievement of discipline through responsible citizenship	4.36	3.91	4.17	3.92
32.	Attention to health conditions	4.28	4.04	4.24	4.06
Mean Ratings		4.66	4.22	4.52	4.21

* Number of Student Teachers in each group

TABLE II
CHARACTERISTICS OF STUDENT TEACHERS CLASSIFIED IN ORDER OF RANK FROM HIGH TO LOW

	Ratings from 2 (low) to 5 (high)			
	Secondary Schools Grades 7-12 Inclusive		Elementary Schools Grades 1-6 Inclusive	
	a	b	c	d
	Secondary Student Teachers Creative Ability	All Secondary Student Teachers	Elementary Student Teachers Creative Ability	All Elementary Student Teachers
	158*	589*	87*	266*
1. Creative ability	1	31	1	31
2. Wholesome philosophy of life	2	3	8.5	3
3. Cooperation with colleagues	3	2	6	2
4. Sincere courtesy of manner	4	1	4	1
5. Social adjustment	6	8.5	4	7
6. Professional mindedness	6	5.	7	7
7. Seeks to grow intellectually	6	8.5	2	4
8. Conscientiousness in meeting responsibilities	8	4	10	7
9. Initiative in achieving objectives	9	17	15	23
10. Emotional balance	11	8.5	4	5
11. Resourcefulness in teaching	11	25	16	24.5
12. Skill in planning	11	19	13	13.5
13. Health and health habits	13	8.5	8.5	10
14. Appearance	14	6	12	9
15. Effective teacher-pupil relationship	16	12	19	17
16. Understands content of his field as needed in schools	16	13	17	20.5
17. Understands social values and implications of his field	16	11	14	12
18. Constructive leadership	18.5	21.5	22.5	29
19. Judicious selection and use of instructional materials	18.5	21.5	20	29
20. Adaptability to a variety of teaching situations	20	23	11	15
21. Sees relationships of his field to all fields	21	17	21	11
22. Success in carrying out plans	22	24	18	17
23. Understands young people	23	17	22.5	20.5
24. Skill in evaluation of pupils' growth and achievement	24	14.5	28	26.5
25. Forcefulness without offensive aggressiveness	25.5	28	24	20.5
26. Possesses a broad cultural understanding	25.5	20	29	26.5
27. Attention to supplies and equipment	27	14.5	30	24.5
28. Provision for individual differences	28	27	25.5	17
29. Attention to English skills as related to subject	29	29.5	27	13.5
30. Effectiveness of speech	30	29.5	25.5	20.5
31. Achievement of discipline through responsible citizenship	31	32	32	32
32. Attention to health conditions	32	26	31	29

* Number of Student Teachers in each group

This may be compared to a range of .68 for unselected elementary teachers, from the highest mean score of 4.60 in "courtesy" to the lowest score of 3.92 in "discipline." Mean scores for the "creative" group in all 32 cases are higher than those for the unselected group. The average mean rating of 4.52 for the "creative" elementary teachers was .31 of a point higher than the 4.21 average of mean scores of the unselected group. All of the 87 students except one, in the "creative" group attained a score of 4.60 or more as a final rating in the course of student teaching, and thus received the final mark of "A." Less than one-third as many of the unselected group were graded as high. This was the case despite the fact that the creative group was included as a part of the unselected group.

All of the above shows clearly that the creative group attained higher ratings in all categories than the unselected group and again raises the question whether creativity was the cause of high ratings in the remaining 31 traits or whether superior student teachers rated consistently high in all respects including "creativity." Facts are unavailable to give a precise answer to this question. It may be pointed out, however, that in spite of the fact that .87 students in the "creative" group were also a part of the unselected group of 266, the ratings in "creativity" on the part of the remaining 179 persons of this group were so low that the mean score of this trait placed it next to the lowest of all 32 items in the rating scale for all elementary student teachers. This is indicative that while there may be some "halo" effect in the case of superior student teachers resulting in a tendency to rate each person high in all 32 traits, "creative ability" is still a quality which serves as one of the chief difficulties in attaining distinctive success in teaching.

An examination of the rank order of traits as presented in columns (c) and (d) Table II discloses several striking differences between the two groups of elementary student teachers. The first of these is initiative which appears as number 15 with the creative group and 23 with the unselected group. In "resourcefulness" likewise there is a difference of eight places in rank in favor of the creative group. "Constructive leadership" and "selection and use of instructional materials" are competencies also appearing significantly higher in the rank of items for creative teachers. While these differences in rank order do not reveal cause-effect relationships between the items, it may be surmised that such relationships exist. Two pertinent questions are: does the resourceful person possess more initiative than otherwise would be the case and to what extent does the possession of both of these qualities result in more effective leadership and proficiency in selection and use of materials? The fact that all four of these items were ranked markedly higher on the list of creative teachers is suggestive that a close relationship exists among them and that they are components of creativeness as defined in this study.

The fact that there are ten places difference in favor of the unselected group in "sees relationships of his field to all fields" is difficult to explain. It may have been caused by the fact that the creative teacher gave his full attention to interpreting and applying the material at hand and less attention proportionately than the less creative teachers to making more formal comparisons with other school subjects.

While mean scores for the creative group are higher for the unselected group in "attention to English skills," "attention to supplies and equipment" and in "provision for individual differences," it may be noted that these three items appear somewhat lower in rank with creative teachers. While data do not permit an explanation, the relatively lower rank may have been caused by proportionately more emphasis being placed by creative teachers on activities they held to be less routine in nature.

Summary and Conclusions

The entire enrollment of 855 student teachers in one year at the University of California, Los Angeles were found as a group to rate next to the lowest of 32 traits in creativeness. Of this enrollment, however, 87 of the 266 elementary student teachers and 158 of the 589 secondary student teachers were rated superior in this quality. In this study, comparisons were made between the highly creative students and the 855 unselected students on the basis of mean scores attained on all 32 traits.

Mean scores for the creative group were higher on all 32 items than was the case with the unselected group of 855. While the ranges of scores of the two groups were not widely different, the range for the creative group was much higher on the scale and thus the averages of mean scores of this group were likewise higher. Because of the fact, therefore, that students who were rated superior in creativeness were also rated high in all other qualities question was raised as to what extent superior creativeness was a cause or result of general superiority in teaching. Because of the fact that (1) supervisors had stressed creativeness as a key point to distinguish between the strong and brilliant teacher (2) ratings for the unselected group in creativeness were so low that its mean score was next to the lowest of all 32 traits and (3) the close relationship between high mean scores in creativeness and general superiority in teaching, the conclusion seems justified that creativeness is essential as a contributor to superior teaching success and that it is proportionately lacking with teachers of inferior ability.

The rank order of the 32 traits disclosed significant differences between the creative and less creative groups. Chief among these were "resourcefulness" and "initiative" which characteristics appeared much higher in rank with both elementary and secondary creative teachers than they did on the list of the unselected group. While the rank order of traits for both creative elementary and secondary

groups were similar there were some noticeable although not significant differences.

Since creativeness appears to be one of the distinguishing differences between the fair and the outstanding teacher, it seems appropriate for teacher education institutions to lay heavy stress on this factor in the process of preparing teachers.

FOOTNOTES

1. Bond, Jesse A. "Strengths and weaknesses of

Student Teachers," Journal of Educational Research, (September 1951).

2. Bond, Jesse A. "Analysis of Observed Traits of Teachers Rated Superior in School Discipline," Journal of Educational Research (March, 1952); "Analysis of Observed Traits of Teachers Rated Superior in Using Speech as a Teaching Instrument," Journal of Educational Research (May 1958).

AN OPINIONAIRE ON WHY COLLEGE STUDENTS CHOOSE TO TEACH

J. MARC JANTZEN
College of the Pacific

A critical shortage of teachers became apparent following World War II. Although a number of factors contributed to this situation, the decreased birth rate during the depression years of the 1930's with its resulting decline in the supply from which to recruit potential teachers, and the increased birth rate during the 1940's with a growth of elementary school enrollments six years later have doubtless been primarily responsible for our great need for teachers.

The current shortage of teachers is stimulating many individuals as well as countless organizations and groups to try to recruit young people for the teaching profession. Increasingly one hears and reads of the term "selective recruitment". In order to develop or improve recruitment techniques it becomes necessary to try to find out why students now in teacher education programs in college have chosen to become teachers. This article compares the results of three investigations, the first of which was conducted in 1946¹, and repeated in 1948 and in 1956² to see if differences or similarities exist during a ten-year span in the opinions expressed by college students on why they had chosen to enter the teaching profession.

In trying to secure information on the question "Why do young men and women choose to teach?" a teacher recruitment committee prepared a check list of sixteen statements of possible factors which might have influenced college students to choose teaching as a profession. This check list, prepared by the committee³ in 1946, was administered to 177 men and 108 women students in the College of the Pacific, Chico State College, and the College of Agriculture of the University of California. In 1948 it was given to 240 men and 227 women students in the College of the Pacific, Chico State College, and Sacramento State College. The 1956 administration was confined to the College of the Pacific with 103 men and 123 women students involved. In each case the sixteen statements were used in an opinion sur-

vey of full-time college students actually enrolled in teacher education courses. The students were asked to check all of the statements that might have influenced them in their choice of becoming teachers. The lists were returned unsigned to avoid any suggestion that the responses would be used to judge individual motives.

As indicated in Table I, the responses to the check list are fairly equally divided between the two sexes, with 520 men and 458 women students. A chi square ($X^2 = 15.966$), indicates a significant difference in the ratios of men and women for the three investigations pointing to the necessity of treating the two sexes separately rather than as a single group.

For purposes of this report, each of the sixteen items of the check list has been condensed into a brief, self-explanatory phrase as it appears in Table II. The statements of the opinion survey are arranged in this table in order of their percentage rank of the 1956 replies of 103 men. A chi square for both men and women indicates no significant variation of ratios of students to the number of questions which they checked in the three studies.

The results of the three studies are summarized in Table II. The per cent of individuals checking each item of influencing factors is presented for each of the surveys made in 1946, 1948, and 1956 with differentiations for men and women.

For example, 68 per cent of the men checked "Interest in children and young people" in 1946 as a reason for choosing to become a teacher, 74 per cent in 1948, and 77 per cent in 1956. For the same item 90 per cent of the women marked this in 1946, 80 per cent in 1948, and 93 per cent in 1956. This factor was given the highest rating by both men and women in all three studies. There was no significant shift for either men or women from one study to the next.

Three other items in the check list remained consistently high for both men and women. These

*Footnotes will be found at the end of this article.

TABLE I
THE NUMBER OF STUDENTS RESPONDING TO AN OPINION SURVEY ON
WHY COLLEGE STUDENTS CHOOSE TO TEACH IN EACH OF THE
STUDIES OF 1946, 1948, AND 1956

	1946	1948	1956	Total
Men	177	240	103	520
Women	108	227	123	458
Total	285	467	226	978

TABLE II
PER CENT ORDER OF REASONS FOR CHOICE OF TEACHING AS INDICATED BY
520 COLLEGE MEN AND 458 COLLEGE WOMEN IN 1946, 1948, AND 1956

Influencing Factors Chosen	Men			Women		
	1946	1948	1956	1946	1948	1956
1. Interest in children and young people	68	74	77	90	80	93
2. Reasonable assurance of adequate income	63	62	73	74	67	71
3. Lifelong opportunity to learn	62	62	61	51	45	63
4. Summer for study, travel, relaxation	67	61	59	74	67	58
5. Enthusiasm of former teacher	36	40	53	40	40	55
6. Specialized training; teach or related business	48	38	48	51	30	29
7. Opportunity for individual initiative	48	43	46	51	45	57
8. Possibilities for advancement	30	38	40	12	24	27
9. Service to mankind	32	31	37	22	31	44
10. Ethics of teaching profession	36	30	36	32	24	39
11. Tenure laws	30	23	35	20	13	19
12. Retirement system	23	18	35	7	9	19
13. Ease of getting position	10	15	24	18	19	19
14. Obligation to society	11	18	17	18	15	34
15. Parents wanted me to be a teacher	3	4	6	19	17	11
16. Family members are teachers	8	14	1	22	20	15

included "Reasonable assurance of adequate income"; "Lifelong opportunity to learn," and "Summer for study, travel, relaxation." Of these three there is a significant shift beyond the .01 level of confidence for women from 1946 to 1956 with reference to an increased checking of the item concerning a lifelong opportunity to learn.

Time has made little change in the degree to which such reasons as "Opportunity for individual initiative", checked by one in two students, and "Ethics of teaching profession", checked by one in three, are given recognition. It is also strikingly apparent the lack of importance attached by these college students to the two items, "Parents wanted me to be a teacher" and "Family members are teachers" as reasons for respondents to this opinionnaire to become teachers.

Marked differences between percentages of men and women are seen in the item of choice indicating "Specialized training: teach or related business" in which men remained at the 48 per cent position in the 1946 and 1956 studies, while women dropped sharply from 51 per cent in 1946 to 29 per cent in 1956. (It should probably be recalled that there were some changes in the institutions participating in the three different surveys.) For the women this represents a significant shift (beyond the .01 level of confidence) apparently indicating that women, once having chosen to become teachers, regard this choice as fairly final as far as professional pursuits are concerned. Men in the 1956 study have checked the item dealing with "Retirement systems" much more frequently than in the previous two studies, a significant shift. Women, on the other hand, continue to consider this item relatively unimportant. They, however, checked "Obligation to society" in the 1956 study twice as often as men.

The 1956 study seems to bear out H. L. Menckens's statement, "A teacher is one who, in his youth, admired teachers." The percentage of men checking "Enthusiasm of a former teacher" increased from 36 per cent in 1946 to 53 per cent in 1956, significant at the two per cent level, with women showing an increase from 40 per cent in 1946 to 55 per cent in 1956, significant at the one per cent level. Two other items checked with considerably more frequency in 1956 than in 1946 are "Possibilities for advancement", showing increases from 30 per cent to 40 per cent for men, and from 12 per cent to 27 per cent for women; and "Service to mankind", increasing from 32 per cent to 37 per cent for men and from 22 per cent to 44 per cent for women. While this shift for the men is not yet a significant one, for the women it represents significance at the one per cent level.

In Table III the chi squares of selected items of the opinionnaire survey are presented to see if any significant shifts from 1946 to 1956 have occurred in the opinions held by college students on why they had chosen to become teachers.

A variation in the 1956 study over those of 1946

and 1948 was a request of the 226 college students to indicate at what point in their academic training they reached the decision to prepare for a teaching career. These data are presented in Table IV. Rather definite sex variations are apparent. Whereas 54 per cent of the women replied that they had decided to teach before they graduated from high school, only 24 per cent of the men had done so. During the first two years of college, 36 per cent additional of the women chose to enter teaching as compared to 50 per cent of the men. Only 10 per cent of the women made their choice to teach as upper division and graduate students, while 26 per cent of the men deferred their decision until then.

If further research substantiates these responses relative to the academic level at which the choice to teach tends to be made, then recruitment efforts should vary for the two sexes. For the girls, recruitment should be rather intensive in high school and extend through the freshman and sophomore years of college; while for the boys recruitment should not be completely ignored in high school, but intensified in the lower division of college and extended into the upper division and graduate years. Apparently teaching as a career is more often a "first choice" with women than with men. Further investigation needs to be done to see if for men it commonly is a second or even third choice. It appears that the final decision to teach is reached much later in the academic years for men than for women.

Conclusions

1. College students who plan to teach maintain consistently over a ten-year period that their number one reason is an interest in children and young people.

2. College students also rate high the assurance of an adequate income and the possibility to continue their intellectual growth.

3. Having the summer for study, travel and vacations is still ranked important but less so in 1956 than in 1946 and 1948.

4. Classroom teachers should be aware of the fact that their own enthusiasm shown on the job may influence young people later to choose to become teachers.

5. Significant shifts in student opinions on why they had chosen to become teachers are indicated for women in checking more often the lifelong opportunity to learn; for both men and women checking more often as an influencing factor in choice for teaching the enthusiasm of a former teacher; for women the possible service to mankind; and for men a teacher retirement system is an increased influence.

6. Women make their choice to become teachers at an earlier age than do men. More than half of the women in the 1956 study (54%) had supposedly reached this decision before they entered college, while only one fourth of the men (24%) indicated do-

TABLE III
CHI SQUARES FOR THE FOLLOWING ITEMS OF TABLE II INDICATING ANY
SIGNIFICANT SHIFTS IN STUDENTS OPINIONS FROM 1946 TO 1956

	Men	Women
1. Interest in children and young people	$X^2=3.209$, not significant	
2. Reasonable assurance of adequate income	$X^2=5.464$, not significant	
3. Lifelong opportunity to learn		$X^2=9.971$, significant at 1% level
4. Summer for study, travel, relaxation	$X^2=2.444$, not significant	$X^2=6.992$, significant at 5% level
5. Enthusiasm of former teacher	$X^2=8.312$, significant at 2% level	$X^2=9.249$, significant at 1% level
6. Specialized training; teach or related business	$X^2=3.620$, not significant	$X^2=16.402$, significant at 1% level
7. Opportunity for individual initiative		$X^2=4.682$, not significant
8. Possibilities for advancement	$X^2=3.827$, not significant	
9. Service to mankind	$X^2=1.217$, not significant	$X^2=12.807$, significant at 1% level
10. Ethics of teaching profession		1946 to 1956: $X^2=1.093$, not significant 1948 to 1956: $X^2=8.961$, significant at 1% level
11. Tenure laws	$X^2=4.329$, not significant	
12. Retirement system	$X^2=11.782$, significant at 1% level	

TABLE IV

THE ACADEMIC LEVEL AT WHICH 226 COLLEGE STUDENTS REPORTED MAKING
THEIR CHOICE TO BECOME TEACHERS WITH SEX
DIFFERENTIATIONS OF CHOICES

Choice made while in	Per cent of 103 men	Per cent of 123 women
Elementary school	1	11
Junior High School	2	4
Senior High School	21	39
College Freshman	26	14
College Sophomore	24	22
College Junior	11	7
College Senior	3	1
Graduate Work	12	2

ing so.

7. Half of the men in the 1956 study chose to become teachers while freshmen and sophomores in college (50%). This would seem to suggest that the junior colleges and lower division of four-year colleges are significant periods during which to recruit men for teaching.

FOOTNOTES

1. Jantzen, J. Marc, "Why College Students Choose to Teach", Phi Delta Kappan, April, 1947, pp. 333-335.
2. Jantzen, J. Marc, Unpublished studies.
3. Members of a teacher recruitment committee

of Xi Field Chapter, Phi Delta Kappa, and their professional positions in 1946 include the following: Joel A. Burkman, Assistant Director of Education, California State Department of Education; Malcolm Murphy, Principal of the Sacramento (California) Senior High School; S. S. Sutherland, Supervisor of Agriculture Teacher Training, College of Agriculture of the University of California; Fred W. Traner, Dean of the School of Education, University of Nevada; Guy A. West, Director of the Division of Teacher Education, Chico State College; and J. Marc Jantzen, Dean of the School of Education, College of the Pacific, chairman of the committee.

THE FIELD TRIP AS A SUPPLEMENT TO TEACHING: AN EXPERIMENTAL STUDY*

JOSEPH S. HILLSON, ALEXANDER A. WYLIE¹
WOLF P. WOLFENBERGER²

In recent years, education institutions of all sorts, and especially Teachers Colleges have made increasing use of a variety of non-academic techniques to supplement the regular classroom experience. Under the philosophy of "active learning" the field trip has particularly seemed to gain in popularity among some educators. The optimistic assumption has been that such experiences are, in fact, educational, although there has been little research justification for such optimism. Since education seems now to be attempting to move from philosophical "schools" to development as a research based science, it no longer seems necessary that we adopt such optimism "on faith" but rather it seems possible, and highly desirable, that we submit such assumptions to the impartial scrutiny which is possible through the application of tried and tested methods of scientific investigation.

The present study was an attempt to test the hypothesis that a field trip is, in fact, an "educational" experience. Specifically, an attempt was made to study the effects of a field trip to a state mental hospital upon prevailing attitudes towards mental hospitals, the field trip being a teaching supplement for an initial college course in general psychology.

Method

Subjects

Both experimental and control groups were composed of undergraduate students currently enrolled in different sections of a required course in general psychology at a teachers college. Since students were assigned to the various sections on a random basis, it seemed a fair assumption that the groups could be considered as drawn from a common population.³ The control group consisted of 18 subjects (8 males and 10 females) while the experimental group consisted of 17 subjects (10 males and 7 females). A previous study (1) has failed to demonstrate any sex differences regarding these attitudes measured in this manner.

Procedure

An attitude scale developed by Souelem (3) was administered to both groups on the same date, under identical conditions.

Three days later the experimental group was taken on a field trip to a state hospital. At the hospital they toured the admission service, the intensive treatment services, recreational and occupational therapy departments, several of the wards, laboratories, professional staff offices, etc. Following this tour of the physical plant they met with the Clinical Director and the Chief Psychologist of the hospital, who presented brief lectures concerning many aspects of the hospital. Following these lectures a lively discussion and question period ensued.

Three days after the field trip both groups were again administered the attitude scale under identical conditions. Prior to the original and follow-up administrations of the attitude scale there was no forewarning to the subjects that they would be tested or re-tested. During the actual administration of the scale, no opportunity was given to the subject's to discuss the scale. Only the items rated for agreement were scored and each subject's attitude score was then represented for further statistical study by the mean score of the various scorable items in his protocol.

Results

Comparison of Groups on the Total Scale

In Table I, the group means and standard deviations of the scale scores obtained in each administration are presented. The higher scores represent more unfavorable attitudes toward mental hospitals whereas the lower scores represent more favorable ones. This, it should be noted, is only relative since the neutral point of the scale is 5.5. The degree to which these group differences are significant is revealed in Table II.

The data demonstrate that the Control Group and Experimental Group did not differ significantly from

* All footnotes will be found at the end of the article.

TABLE I
MEANS AND STANDARD DEVIATIONS OF GROUP ATTITUDE SCALE SCORES

	Control Group		Experimental Group	
	First Administration	Second Administration	First Administration	Second Administration
Number	18	18	17	17
Mean	3.65	3.69	3.68	3.38
S. D.	0.35	0.30	0.26	0.32

TABLE II
TESTS OF THE SIGNIFICANCE OF THE DIFFERENCE
BETWEEN GROUP MEAN SCALE SCORES

	Difference	S. E. Difference	<u>t</u>	Significance Level
Control ₁ - Experimental ₁	0.03	.0255	1.18	n. s.
Control ₁ - Control ₂	0.04	.0260	1.54	n. s.
Experimental ₁ - Experimental ₂	0.30	.0250	12.00	.001
Control ₂ - Experimental ₂	0.31	.0255	12.16	.001

TABLE III
SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE MEAN SCORE FOR EACH SUBJECT
OF THE EXPERIMENTAL GROUP IN FIRST AND SECOND ADMINISTRATIONS

Subject	First Administration	Second Administration	Differ- ence	S. E. Difference	t	Significance Level
1	4.33	3.67	0.66	0.087	7.59	.001
2	3.33	3.32	0.01	0.	0.	n. s.
3	3.59	3.24	0.35	0.089	3.93	.001
4	3.57	3.25	0.32	0.091	3.51	.001
5	3.55	3.05	0.50	0.071	7.04	.001
6	4.03	3.99	0.04	0.	0.	n. s.
7	3.45	3.10	0.35	0.083	4.22	.001
8	3.88	3.71	0.17	0.089	1.91	n. s.
9	3.82	3.58	0.24	0.080	3.00	.01
10	3.70	3.23	0.47	0.074	6.35	.001
11	3.90	3.72	0.18	0.071	2.54	.05
12	3.55	3.23	0.32	0.084	3.81	.001
13	3.20	3.46	0.26	0.084	3.10	.01
14	3.39	2.79	0.60	0.099	6.06	.001
15	4.09	3.71	0.38	0.100	3.80	.001
16	3.83	3.15	0.68	0.078	8.71	.001
17	3.29	3.27	0.02	0	0	n. s.



each other at the time of the first administration regarding the attitudes measured by the scale. This indicates that the Control Group and Experimental Group were, in fact, drawn from a common population.

The results of the second administration of the attitude scale revealed that the overall attitude of the Control Group has remained constant, while that of the Experimental Group has undergone modification in the positive direction to a degree which is significant beyond the .001 level of probability. Likewise, the attitude held by the Experimental Group now differed from its earlier attitude to an extent which was equally significant.

Since the test-retest situation with the Control Group was, in effect, a measure of reliability, a coefficient of reliability was computed. The obtained r of .92 indicates that the attitude scale has adequate reliability.

Intra-Subject Comparisons of the Experimental Group

Since it seemed desirable to ascertain whether the shift in the group score for the experimental group was the result of a general trend or the result of marked changes in only a few subjects, a test-retest comparison for each individual's performance in this group was undertaken. The data for this comparison are presented in Table III. These data indicate that 13 of the 17 experimental subjects modified their response in the second condition to a degree which was highly significant. It would seem that the experimental condition had an effect upon approximately three-fourths of the group. All of these changes were in the direction of a shift toward a more favorable attitude with the exception of subject No. 13, who changed his attitude in the unfavorable direction. It is noteworthy that in the original administration, this subject responded in a manner such as to indicate that he held the most favorable attitude toward mental hospitals of the entire group. The experience of the field trip seems to have shifted his score much closer to the mean for the entire group.

It was also noted that three of the remaining four subjects who did not show a significant shift, deviated from the mean of the group in the original ad-

ministration by more than one standard deviation.

Summary and Conclusions

An attempt was made to measure the effects of a field trip experience upon the prevailing attitudes of a group of college students toward mental hospitals through the use of a standardized attitude scale. The experimental condition was found to produce a small, but highly significant, shift in the attitudes of the group in the direction of a more favorable attitude. The fruitfulness of such an approach seems to have been demonstrated and further investigation may delineate the manner in which such experiences can be made more effective.

Since neither the generality nor the stability of the observed results have yet been demonstrated, the authors would urge due caution in generalization from the present study.

FOOTNOTES

- * The authors wish to express their gratitude to W. A. Brandenburg, Ph.D., President, Nebraska State Teachers College (Wayne), and to C. G. Ingham, M.D., Superintendent, Norfolk (Nebraska) State Hospital, at which institutions this study was performed, for their fine cooperation and for the assistance offered by many other members of both institutions.
- 1. Norfolk (Nebraska) State Hospital and Nebraska State Teachers College (Wayne).
- 2. George Peabody College for Teachers.
- 3. The experimental design included a test of this assumption and it was found to be justified.

BIBLIOGRAPHY

1. Klopfer, W. G., Wylie, A. A., Hillson, J. S., "Attitudes Toward Mental Hospitals," Journal of Clinical Psychology, XII, No. 4, (1956), pp. 361-365.
2. Soulelem, Omneya "Mental Patient's Attitudes Toward Mental Hospitals" Journal of Clinical Psychology, XI (1955), pp. 181-185.
3. Soulelem, Omneya, A Scale Measuring Attitudes Toward Mental Hospitals, Unpublished M. A. Thesis, Graduate School of the University of Oregon, Eugene, Oregon.

A STUDY OF TWENTY SLOW LEARNERS

SISTER AGNES JEROME
University of Hawaii

THE PROBLEM of the slow learner has always been present in education. It is only in this century, however, that any serious and successful attempt has been made to measure the range of individual differences. This recent endeavor has succeeded in upsetting many previous assumptions about the possible progress of individual students through a graded school.

The Problem

The assumption was made, on the findings of investigators in the fields of Psychology and Education, that learning is significantly impeded not only because of retardation in intelligence, but also because of such factors as socioeconomic status, personality traits, physical defects and classroom adjustment. It was the belief of the present investigator that one or more of these factors will be found in the learning patterns of slow learners. More specifically, this study was concerned with the following question: What patterns of factors prevalent in slow learning children of several classes in one school can be discovered by an intensive study of such cases?

Subjects

The subjects of this study were selected from 500 cases. They were the twenty children in one school having the lowest educational quotient in grades 3 to 8. They were classified as "slow learners" under the following conditions:

1. If their E. Q. (Educational Quotient) was considerably below average.

2. If they were achieving one or more years below their present grade. On September 1, 1957, the range in chronological age of these children was 9 years, 5 months to 14 years, 2 months. The mental ages ranged from 8 years, 4 months to 13 years, 4 months. Intelligence quotients, as measured by the Revised Stanford-Binet, ranged from I. Q. 74 to I. Q. 110, the mean I. Q. being 89.

The sources of information used in the study

were:

1. S. R. A. Achievement Tests.
2. Results of the Revised Stanford-Binet Tests of intelligence, Form L.
3. School records and files.
4. California Test of Personality, Form AA, Elementary Series.
5. Behavior Preference Record, Form A.
6. Mental Health Analysis, Elementary Series.
7. Teachers' Rating Scale.
8. Vineland Social Maturity Scale.
9. Health Records.
10. Personal interviews with all individuals in the study.
11. Interviews with parents and teachers of the individuals in the study.
12. Bonney-Fessenden Sociograph.

From these sources the following 15 factors were selected to secure the data utilized in the study:

1. Educational quotient.
2. Length of time in attendance at this school.
3. Reading quotient.
4. Spelling quotient.
5. Arithmetic quotient.
6. Achievement quotient.
7. Chronological age.
8. Binet mental age.
9. Binet intelligence quotient.
10. Profiles of personality traits.
11. Sociogram indicating Peer Status.
12. Social quotient.
13. Teacher rating.
14. Estimate of socioeconomic status.
15. Summary of physical factors.

From this data individual case studies were written and analysis made.

Findings

An analysis was made of the data collected and findings were reported under four categories:

I. Intellectual Findings. The data collected indicated a definite degree of relationship between re-

TABLE I
COMPARISON OF I. Q. , E. Q. AND S. Q.

Case	Chronological Age	Intelligence Quotient (Binet)*	Educational Quotient (S. R. A.)**	Social Quotient (Vineland)***
1.	12.10	81	73	81
2.	12.7	82	77	95
3.	13.1	84	80	84
4.	13.7	93	83	114
5.	13.7	91	84	100
6.	14.1	95	84	104
7.	14.2	74	84	105
8.	13.1	102	85	96
9.	12.9	90	86	96
10.	13.2	90	88	98
11.	13.7	97	88	110
12.	13.10	82	89	102
13.	14.0	89	90	103
14.	14.0	84	91	107
15.	9.5	97	91	109
16.	11.1	110	92	91
17.	12.5	82	92	100
18.	10.8	86	94	91
19.	11.0	76	95	99
20.	11.0	97	96	100

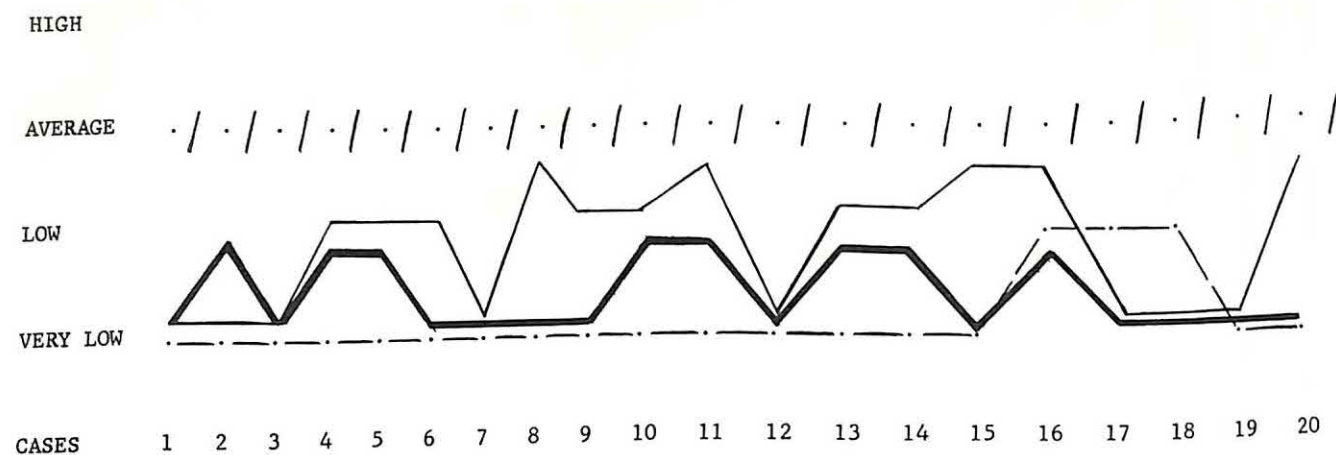
* Stanford-Binet Test
** Science Research Associates Ach. Test
*** Vineland Social Maturity Test

TABLE II
RANKING ON ALL ADJUSTMENT INSTRUMENTS

Case	California Test of Personality	Behavior Preference Record	Mental Health Analysis	Bonney- Fessenden Sociograph	Vineland Social Maturity	Teachers' Rating Scale	Total
1.	2	2	2	2	1	1	1
2.	2	2	1	3	2	1	2
3.	2	2	1	3	1	1	1
4.	4	3	1	1	4	3	2
5.	2	2	3	2	3	1	2
6.	1	2	1	1	3	1	1
7.	2	1	2	1	3	1	1
8.	2	1	1	2	2	1	1
9.	2	2	1	1	2	2	1
10.	3	2	1	2	3	2	2
11.	2	2	3	1	4	1	2
12.	2	1	1	2	3	1	1
13.	3	3	3	1	3	1	2
14.	2	2	2	1	4	3	2
15.	1	2	1	1	4	1	1
16.	1	3	2	2	2	1	2
17.	2	2	1	1	3	1	1
18.	1	1	1	2	2	1	1
19.	1	1	2	1	3	1	1
20.	1	2	1	1	3	1	1

Very Low
Low
Average
High

-1
-2
-3
-4

FIGURE 1.Legend

Intellectual ———
 Academic - - - - -
 Adjustment ———
 Physical . / . / .

tardation and low intelligence among the twenty cases studied. Sixty per cent of the subjects had I. Q.'s ranging from 74 to 90. The mean for the entire group was 89. Because of limited mental capacity, satisfactory performance in the tool subjects was impossible for these cases.

II. Academic Findings: The hypothesis that a relationship exists between retardation and failure in the tool subjects found support in the study. According to the results of the standardized tests (S. R. A.), 100 per cent of the cases were very low in reading comprehension; fifty per cent were below average in spelling; ninety per cent were below average in arithmetic.

III. Adjustment Findings: Total adjustment data through evaluation of the six adjustment instruments afforded a broad base for adequate appraisal of pupil adjustment in that three types of measurement were secured. The peer estimate and pupil self-ratings served as a balance and check on the teachers' assessment, which for the most part was based on personal observation. Table II shows that twelve cases or sixty per cent of the subjects studied were very low in total adjustment and that the remaining forty per cent were low.

IV. Physical Findings: Research has indicated that there is some degree of relationship between retardation and such physical factors as absenteeism, general health, visual defects, auditory defects, and motor coordination. In this study, however, it was found that all cases were in satisfactory physical condition at the present time. Two cases had reported minor physical disturbances in the past but no chronic conditions that might in any way inhibit learning.

V. Findings Related to the Patterns: Figure 1 shows graphically the interrelationship among the four factors used as a basis in this investigation. A study of the individual profiles revealed eight patterns. Two types of curves were evident in these patterns--the "reversed L" and the "U curve".

3. Twelve cases showed achievement lower than ability; their profiles were U-shaped curves, with academic achievement and adjustment lower than intelligence.

4. All cases showed ratings as average on physical findings.

5. The patterns as a whole indicate three levels of intelligence; three groups score very low, two groups score low, and three show average intelligence.

6. Six of the eight patterns indicate very low adjustment.

REFERENCES

1. Barbe, Walter H. "Peer Relationships of Children of Different Intelligence Levels," School and Society, 80: 334-336, June, 1954.
2. Baron, Denis M. Personal Social Characteristics and Classroom Social Status, Unpublished Master's thesis, the University of Oregon, 1951.
3. Bronner, Augusta F. The Psychology of Special Abilities and Disabilities, Boston: Little and Brown, 1947, p. 37.
4. Heck, A. O. "An Analysis of Problems Involved in the Study of School Failures," Educational Research Bulletin, (Ohio State University), 1927, 6: 67-72.
5. Powell, H. F. Social Adjustment as Related to Reading Disability, Unpublished Master's thesis, the University of Michigan, 1943.
6. Reeder, Thelma A. A Study of Some Relationships Between Level of Self-Concept, Academic Achievement, and Classroom Adjustment, Unpublished Doctor's thesis, North Texas State College, 1955.
7. Spache, George. Personality Patterns of Retarded Readers, Unpublished Master's thesis, the University of Florida, 1955.

		Intellectual	Academic	Adjustment	Physical	No. Cases
Pattern I		Very low	Very low	Very low	Average	5
Pattern II		Very low	Very low	Low	Average	1
Pattern III		Very low	Low	Very low	Average	2
Pattern IV		Low	Very low	Very low	Average	2
Pattern V		Low	Very low	Low	Average	5
Pattern VI		Average	Very low	Very low	Average	2
Pattern VII		Average	Very low	Low	Average	2
Pattern VIII		Average	Low	Low	Average	2

An analysis of the data presented on these patterns may be summarized as follows:

1. Six subjects, with I. Q.'s of 81, 82, 83, 74, 82, 76, were achieving approximately at ability; their profiles were reversed L's with very low rankings on intelligence, achievement and adjustment.

2. Two cases were over-achievers; their profiles were irregular with achievement higher than intelligence and adjustment.

8. Torgerson, Theodore L. Studying Children, New York: The Dryden Press, 1947, pp. 32-38.
9. Tulchin, Simon H. "Emotional Factors in Reading Disabilities of School Children," Journal of Educational Psychology, 26: 23-26, 1935.
10. Woody, Clifford. "An Analysis of Differences in the Learning of Bright and Dull Children," School Bulletin (University of Michigan), 6: 67-72, 1936.

SOME PERCEPTUAL CHANGES DURING SENSITIVITY TRAINING

KAJ LOHMANN, JOHN H. ZENGER, IRVING R. WESCHLER
University of California, Los Angeles*

THIS EXPLORATORY study was conducted to determine whether changes occur in students' self-perceptions and their perception of trainers during a sensitivity training experience.¹

Sensitivity training--a form of human relations training--is designed to help people become more effective in dealing with others, both on an individual and on a group basis. The underlying objectives include the development of social sensitivity, i. e., the ability to sense how others think and feel, and of behavioral flexibility, i. e., the ability to act appropriately on the basis of these understandings. More specific goals are the development of greater self-understanding, insight into interpersonal relations, recognition of organizational and cultural variables, and improved communication skills.

Method

The subjects were male and female college students in three sensitivity training classes, entitled "Leadership Principles and Practice," meeting twice a week, in two-hour sessions, for sixteen weeks. The three classes had 24, 22, and 19 students respectively and were conducted by three male trainers. All sections were characterized by relative lack of direction and structure. Certain limitations were imposed on the groups, such as the use of final grades, term projects, and examinations. Within these limits, the classes were relatively free to function as they saw fit. The trainers occasionally injected comments and suggestions, but primarily acted as resource persons. Though content material was seldom stressed in class discussions, selected readings were assigned by the trainers to give the trainees a conceptual framework for their experiences.

The Gordon Personal Profile (1) was administered to all students near the beginning and at the end of the training experience. In both instances, the students first completed the form with reference to themselves, providing the student self-perception scores; they then repeated the process with reference to how they saw their trainer, yielding the

trainer-perception scores. The trainers also completed the inventory with reference to themselves, once midway in the training process, yielding the trainer self-perception scores.

The Gordon Personal Profile is designed to yield measures of five aspects of personality: Ascendancy (A), Responsibility (R), Emotional Stability (E), Sociability (S), and Total Self-Evaluation (T). High scorers on the first four scales are described by the manual as follows: With regard to scale A, they are individuals who adopt an active role in group situations, who are self-assured and assertive in relationships with others, and who tend to make independent decisions.

With regard to scale R, they are individuals who take responsibilities seriously, who are able to stick to any job and get it done, and who are persevering and determined. Scale E characterizes high scorers as well-balanced, emotionally stable individuals who are relatively free from anxiety and nervous tension, while scale S identifies them as individuals who like to be with and work with people who are gregarious and sociable.

The total (T) score is obtained by algebraically adding the scores on the four scales (A, R, E, and S). The possible range of scores is from -18 to +18 for the individual scale scores and from -36 to +36 for the Total scores. High T scores indicate that the respondents have marked most of the items in a manner generally considered complimentary to themselves. Although such scores are probably influenced by the tendency of many individuals to overrate themselves on self-description tests of personality, in most instances they probably reflect genuine feelings of confidence and adequacy.

Hypotheses, Findings, and Discussion

1. It was predicted that the trainers would be seen by the students as more adequate at the beginning of training than at the end, i. e., increased acquaintance in a permissive setting was expected to lead to diminished idolization. (The findings of this study are summarized in Table I.)

* Footnotes will be found at the end of the article.

TABLE I
GORDON PERSONAL PROFILE SCORES AT BEGINNING AND END OF TRAINING

		Ascendency		Responsibility		Emotional Stability		Sociability		Total	
		Before	After	Before	After	Before	After	Before	After	Before	After
Group 1											
Student self-perception scores	\bar{X}	5.6	5.3	4.8	5.1	4.5	5.2	6.9	6.3	21.9	21.8
	σ_m	5.5	5.9	6.8	6.4	6.4	6.5	4.6	5.5	15.1	14.7
Trainer-perception scores	\bar{X}	9.0	8.0	8.2	9.0	10.2	9.1	7.4	6.0	34.8	**32.0
	σ_m	3.4	3.3	3.7	4.2	3.3	4.5	4.3	3.5	4.8	3.4
Trainer self-perception scores		4		8		4		4		20	
Group 2											
Student self-perception scores	\bar{X}	5.0	5.4	6.2	5.5	5.0	4.6	7.1	6.6	23.4	22.5
	σ_m	5.6	6.8	4.5	5.4	4.4	6.9	5.9	7.1	6.0	15.7
Trainer-perception scores	\bar{X}	9.8	10.0	5.6	5.5	9.0	7.2	7.2	7.2	31.7	30.4
	σ_m	3.7	4.4	3.6	4.2	3.8	4.8	3.7	4.3	1.4	6.4
Trainer self-perception scores		6		7		15		2		30	
Group 3											
Student self-perception scores	\bar{X}	3.5	5.7	4.2	5.0	0.6	3.0	5.4	5.8	13.7	19.6
	σ_m	7.0	5.9	3.2	5.1	7.6	5.2	7.3	6.0	16.3	15.7
Trainer-perception scores	\bar{X}	8.4	8.3	7.3	6.6	4.7	2.0	8.4	7.7	28.7	24.4
	σ_m	9.1	4.2	4.3	4.9	4.9	6.9	5.6	3.1	10.7	9.0
Trainer self-perception scores		8		7		-2		7		20	

* Test of Significance = 0.05% level of confidence.

** Test of Significance = 0.01% level of confidence.

By comparing the total trainer-perception scores, before and after, we find diminished idolization of the trainers in all cases, the difference being statistically significant for group 1 as well as for all groups lumped together. Such diminished idolization seems a desirable outcome of sensitivity training in that it probably reflects a more realistic perception of the trainers. Although almost all of the initial trainer perception scores seem to reflect stereotyped expectations by the students of what a teacher-trainer should be like, individual differences in idolization among the trainers are apparent. These differences can easily be traced both to the personalities of the three trainers and to the training methods employed by them, as noted by trained observers. Trainer 1 consistently seems to convey personal adequacy as well as confidence of what he is or is not doing in his training role. Trainer 2 remains fairly aloof from his group, at least at the beginning of training, and gives off few cues concerning his personality. Trainer 3 frequently alternates between active and inactive participation in his group; he thus becomes controversial for his group by failing to offer a stable frame of reference against which the group can judge either his personality or the effectiveness of his training methods.

2. It was predicted that the mean total student self-perception scores in each group would remain lower than their mean total trainer-perception scores.

Such a trend exists in all groups, but the hypothesis was statistically confirmed by a significant difference only in group 1. The tendency for groups to continue to see their trainers as more adequate than themselves, despite diminished idolization, can be explained as a function of the expectations and needs which the trainees have in regard to their trainers. If the trainer is seen by the group as psychologically less adequate (i. e., scoring lower) than most of its members, he is likely to be considered as having little to offer and to be rejected in his training role. Such a group may find itself psychologically leaderless unless effective leadership is able to emerge from within its own ranks.

One interesting exception to the trend occurs in the after-training scores of group 3 with reference to factor E (Emotional Stability). In this instance, the group's mean self-perception score is higher than the group's mean trainer-perception score. When comparing the over-all mean trainer-perception scores of the three trainers, it will be noted that the deficit in the total score for trainer 3 can be accounted for almost entirely by his deficit in factor E. Although the group is not at all sure about the soundness of its judgments with regard to trainer 3--the large standard deviations serve as a measure of the trainer's controversial role--it nevertheless may have perceived trainer 3 realistically in view of his own low self-rating with reference to this particular factor.

3. It was predicted that the mean student self-

perception scores would rise as a result of the training experience.

This hypothesis was not confirmed. Only in group 3 did the students see themselves as more adequate at the conclusion of training and, even there, the findings are not statistically significant.

These results are somewhat surprising in that sensitivity training has at least as one of its objectives the improvement of individuals' feelings of self-adequacy. In this regard, it is similar to group psychotherapy which, of course, places considerably greater stress on the accomplishment of this particular objective.²

How, then, can these results be explained? First, sensitivity training may not cause major personality reorganization. The trainees are, by and large, non-neurotic individuals, whose goals for participating in sensitivity training are much more external than internal--that is, they are more concerned about introducing changes in others than in themselves. Second, the kinds of personality changes which do occur during sensitivity training are different in nature (less deep) than those which the Gordon Personal Profile was originally designed to tap. Third, as a paper-pencil test, the Gordon Personal Profile may not be sensitive enough to pick up major personality changes even if they do occur. Fourth, some of the trainees may have attempted to give socially acceptable answers, especially during the first administration of the test, when they felt insecure about how these test results were to be used. This alternative, however, has little to support it in view of recent findings which suggest that the Gordon Personal Profile is probably less subject to "faking" than other inventories of its type (2).

Finally, the hypothesis itself may be based on faulty assumptions. If sensitivity training succeeds in helping trainees see themselves more realistically, both with reference to their assets and their liabilities, then the newly-gained insights may have changed some of the distorted self-images which the trainees held at the beginning of training. With increased confidence in themselves, the trainees might have learned to accept unpleasant characteristics and behaviors in themselves which their defenses prevented them from facing initially. As a result of seeing themselves "better", the trainees' final responses on the Gordon Personal Profile may reflect changes in both positive and negative directions, thus preventing an increase in over-all self-acceptance scores.

4. It was predicted that the groups' final trainer-perception scores would correspond more closely to the trainers' self-perception scores than did their initial trainer-perception scores.

It was hoped that the trainees would show increased social sensitivity by seeing behind the facades established by their respective trainers. The findings confirm a trend in the hypothesized direction. There is no guarantee that the trainers' self-perception scores are more accurate than the

trainees' perception of them, though we believe the trainers to be more aware of their own needs and defenses by virtue of their professional training and experience. The tendency of the trainees more accurately to approximate the self-perception scores of their respective trainers might therefore be interpreted as partial success of the training to increase social sensitivity.

Summary and Prospects

The Gordon Personal Profile was administered to three classes, both at the beginning and at the end of a sensitivity training experience. Students were asked to complete the form, first with reference to themselves, and second, with reference to how they saw their respective trainers. The trainers also completed the inventory with reference to themselves. The analysis attempted to test various hypotheses concerning perceptual change as revealed through student self-perception, trainer perception, and trainer self-perception scores.

Currently, research is underway to test more intensively the impact of sensitivity training. Among the problems being investigated are the personality dimensions along which trainees differ, both before and after training; the attitudes which they have with reference to their participation in groups, both before and after training; how they are chosen sociometrically on a number of criteria, by the trainers, their peers, and a clinically trained staff observer, both before and after training; how they react in diaries to the training process as it proceeds; how the personalities and interventions of the trainers influence the training process; and finally, in line with present research emphasis, what effect the trainees have on the behavior of their trainers.

FOOTNOTES

- * Human Relations Research Group, Institute of Industrial Relations and School of Business Administration.
- 1. For a description of sensitivity training, see, for example, Irving R. Weschler, Marvin A. Klemes, and Clovis Shepherd, "A New Focus in Executive Training," Advanced Management, 20: 17-22 (1955).
- 2. A number of evaluation studies of psychotherapy in general, and group psychotherapy in particular, have shown changes in the predicted direction of improvement in self-acceptance. Most of these findings, however, are based on content analyses by qualified judges of clients' responses during counseling interviews with regard to a number of relevant dimensions, including self-approval and disapproval, while the authors of this study have used scores on the Gordon Personal Profile as indicators of change. (For a review of some of the major studies evaluating therapy, see Carl R. Rogers and Rosalind F. Dymond, Psychotherapy and Personality Change, Chicago, University of Chicago Press, 1954.)

REFERENCES

1. Gordon, L. V., The Gordon Personal Profile Manual. Yonkers, New York, World Book Company (1953).
2. Rusmore, Jay T., "Fakability of the Gordon Personal Profile," Journal of Applied Psychology, 40: 175-177 (1956).

THE RELATIONSHIP OF AUTHORITARIANISM TO REJECTION

JACK R. FRYMIER
Temple University

The Problem

Are authoritarians rejected by their peers? Meer (8) reported that authoritarians showed more aggression toward outgroup members than ingroup members. Rohde (10) found a negative correlation between acceptability and authoritarianism among 176 aircrew members. Hollander (6) tested 168 Naval pre-flight students and found a negative correlation between authoritarianism and nomination by peers for a leadership position. Medalia (9) studied 298 Air Force men and observed a positive relationship between authoritarianism and leader acceptance. Buswell (2) reported that achievement in reading among first grade children was related to acceptance by members of the peer group. She also felt that reading achievement caused or helped bring about the acceptance. Kinney (7) studied five classes of fifth grade students and reported that it was possible to increase the acceptance of individuals over a period of time by working toward such an objective. In an unpublished study Frymier (4) reported that there was a positive relationship in the way in which individuals were accepted and the length of time they had attended school together.

The Procedure

To determine whether or not authoritarians were rejected by their peers, the author measured a group of high school students in a laboratory school sociometrically for rejection. Seventy-one members of this group were tested later with a measure of authoritarianism. Finally, the relationship between rejection scores and authoritarianism scores was determined by computing the coefficient of correlation.

Specifically, five core sections, grades 9 through 12, were asked to respond to the following questions:

1. List here the two persons you would prefer to sit between in core.
2. List here any persons you would prefer not to sit with in core.
3. List here the two persons you would select to be on a committee studying some unit in core with you.
4. List here any persons you would prefer not to work with on a core committee.

If a student's name was listed in response to question one or three, this was assumed to indicate acceptance. If a student's name was listed in response to question two or four, this was assumed to indicate rejection.

The results for each of the 71 persons involved in this study were tabulated as follows: Each time a student's name was listed after either question one or three, he received one plus score (plus interpreted as acceptance), whereas each time a student's name was listed in response to either question two or four he received one minus score (minus interpreted as rejection). The results of these four questions were added algebraically for each student. Finally, these scores were all converted to positive numbers by assigning the highest acceptance score a value of 1 and thereafter adding one to each possible score of acceptance through rejection. In this way the highest acceptance score +17 became 1, +16 became 2, +15 became 3, +14 became 4, +13 became 5, and so on through zero and the negative (rejection) scores up to -20, the highest rejection score for any member of this particular group. In the procedure thus employed, then, the acceptance-rejection scores, which ranged from +17 to -20, were all converted to rejection scores ranging from 1 to 38, the higher scores indicating greater rejection.

Approximately seven months after they had been measured for rejection, 71 members were tested with the F (Fascist) scale to determine their ten-

1. Questions one and three were the only ones the students were required to answer. It was explained to them that questions number two and four did not necessarily have to be completed unless they felt that way about some particular person(s).

dency toward authoritarianism. Individual F scale scores were computed according to the instructions of Adorno (1). Mean scores were also determined. For the purposes of this study, high F scale scores were assumed to indicate a tendency toward authoritarianism and low F scale scores were assumed to indicate a tendency toward nonauthoritarianism.

The two scores thus determined for each of the 71 persons, rejection and authoritarianism, were compared by computing the coefficient of correlation ("r") between pairs.

Discussion of Results

The members of this particular group had rejection scores ranging from 1 to 38 with a mean value of 16.87. F scale scores ranged from 1.67 to 6.33, with a mean value of 3.99 (as compared to a theoretical mean value of 4.00). The coefficient of correlation was .31 between pairs of scores for members of this group.

These data seem to indicate that authoritarians tend to be rejected by their peers. Authoritarians apparently are perceived as being less desirable in this situation than nonauthoritarians. This would seem to corroborate the findings of Rohde and Hollander.

Such a conclusion, however, is based upon the assumption that the F scale is a valid measure of authoritarianism, and there is some evidence (3, 5, 11) that this may not be so. The results reported here, then, might be interpreted as supporting the logical conclusion that persons who are high F scale scorers will be rejected more than persons who are low F scale scorers in a culture which considers authoritarianism less desirable than nonauthoritarianism. If this is true, these data may mean that the F scale measures something which, if not authoritarianism, is at least socially undesirable.

REFERENCES

1. Adorno, T. W., et. al. The Authoritarian Personality. New York: Harper and Brothers, 1950.
2. Buswell, M. M. "The Relationship Between Social Structure of the Classroom and the Academic Success of the Pupils." Journal of Experimental Education, September, 1953, pp. 37-52.
3. Christie, R. and Jahoda, M. Studies in the Scope and Method of "The Authoritarian Personality". Glencoe: The Free Press, 1954.
4. Frymier, J. R. "Acceptance and Rejection as Related to Length of School Attendance." Unpublished research, Temple University, 1957.
5. Frymier, J. R. "Analysis of Adolescents' Responses to the F Scale." Unpublished research, Temple University, 1958.
6. Hollander, E. P. "Authoritarianism and Leadership Choice in a Military Setting." Journal of Abnormal and Social Psychology, July, 1954, pp. 365-370.
7. Kinney, E. E. "A Study of Peer Group Social Acceptability at the Fifth Grade Level in a Public School." Journal of Educational Research, September, 1953, pp. 57-64.
8. Meer, S. J. "Authoritarian Attitudes and Dreams." Journal of Abnormal and Social Psychology, July, 1955, pp. 74-78.
9. Medalia, N. Z. "Authoritarianism, Leader Acceptance, and Group Cohesion." Journal of Abnormal and Social Psychology, September, 1955, pp. 207-213.
10. Rohde, K. J. "The Relation of Authoritarianism of the Aircrew Member to His Acceptance by the Airplane Commander." American Psychologist, July, 1951, p. 323.
11. Titus, H. E. and Hollander, E. P. "The California F Scale in Psychological Research: 1950-1955." Psychological Bulletin, January, 1957, pp. 47-64.

A COMPARISON OF NORMAL AND EMOTIONALLY ILL CHILDREN ON THE KAHN TEST OF SYMBOL ARRANGEMENT*

HOWARD H. FINK, THEODORE C. KAHN**

The Kahn Test of Symbol Arrangement (KTSA) represents a new approach to personality assessment (5, 1). The materials used in the test have shapes of commonly recognized objects and symbols (3). The person taking the test indicates how he reacts to stimuli which are familiar to him and which have meaning for him (2, 3). This is quite different than responding to ink blots. It is also different from pictorial representation, as in the Thematic Apperception Test or the Blacky pictures where the subject is confronted with a picture where everything is in place and structured in such a way that he cannot make any changes in the stimuli materials (4). Although the KTSA uses objects which are "culturally structured" in the sense that they are familiar to all of us, it permits the subject to structure and manipulate the test situation by requiring that he arrange, discuss and judge the objects (2). He arranges them five times under varying conditions. The total task usually takes no more than fifteen or twenty minutes.

In all, there are sixteen objects, one of which is used only briefly during the test performance. The objects are: three heart-shapes, differing in color, size, thickness and translucency; three stars, two of which are identical; two butterfly-shapes, varying in outline, size, width and color; a green parrot, a blue anchor, a transparent circle, an equilateral cross, and three dogs. Two of the dogs are alike except in color; the third resembles one of the others in color but not in shape. If one wishes to combine the items for their shapes or cultural associations, one must sacrifice logical color combinations. If the color choice is made, the objects cannot be combined for their meaning or their shape, size or thickness.

Subjects are asked to arrange the objects along a felt strip in any way they choose. They are then asked to indicate why they arranged them the way they did. A subsequent task is to ask the subjects

to recall as well as they can how the objects were previously arranged and to guess how many they would get right when they do it again. They are also asked to guess on completion of this task how many they correctly placed. In this way, levels of aspiration and achievement are obtained (4). The subjects are also asked to rank the items in terms of their preference, and they are asked to tell what each object could symbolize for them.

Method

A total of 148 children from public schools in Dayton, Springfield and Fairborn, Ohio, were given the KTSA. The age range of this group was from 6 years to 15 years, 10 months (mean age, 9.9 years). The grade range was from first to eighth grade. Intelligence test scores were available for 77 children in this group. These 148 children were judged to be without emotional illness by their teachers and by the school psychologists. These judgments were based on classroom behavior and general adjustment.

The KTSA was also administered to a group of 49 children seen at the Child Guidance Center for Dayton and Montgomery County. These children were judged to be emotionally disturbed by psychologists and psychiatrists. The age range for the Child Guidance Center group was from 6 years to 16 years, 5 months (mean age, 10.0 years). The grade range was from first grade to ninth grade.

The KTSA was scored according to the manual, and a symbol score was obtained for each child. Correlations were run between the symbol scores and the child's chronological age, and between the symbol score and the child's mental age on those subjects where the MA was available.

Results

The results of the correlations are to be seen in

* Read at Annual Meeting of the American Psychological Association, New York City, 1957.

** Howard H. Fink, Child Guidance Center for Dayton and Montgomery County, Dayton, Ohio, and Theodore C. Kahn, 7100 USAF Hospital, Wiesbaden, Germany.

the following figures:

GROUP I - Normal Children

r_{KC}^1	.53** (N=148)
r_{KM}^1	.46** (N=77)

GROUP II - Emotionally Ill Children

r_{KC}^1	.18* (N=49)
r_{KM}^1	.29* (N=49)

¹K - KTSA Symbol Score
C - Chronological Age
M - Mental Age

* Significant at better than the .05 level
** Significant at better than the .01 level

These results indicate, for the normal group, positive significant correlations between symbol score and chronological age, and between symbol score and mental age. For the emotionally ill group, there was a significant but lower correlation between symbol score and mental age, and no significant correlation between symbol score and chronological age. Other results seen upon comparing the two groups are: (a) Normal children had higher test scores. (b) Normal children were able to make higher level associations. (c) Normal children were capable of better abstractions. (d) Normal children had a greater frequency of acceptable social interpretations of the meaning of the objects. (e) Disturbed children typically used a naming approach to the task of associating; i. e., instead of saying what the objects represented or stood for, they merely named what they were, calling a heart a heart instead of saying "it stands for love". (f) Emotionally ill children did not react to cultural symbols in the same manner as did the normal children. They more frequently rejected the symbol and showed an unwillingness to use them constructively. (g) Disturbed children dealt with the symbols and obtained test scores in a manner more nearly like the performance of normal children three years younger than the disturbed children.

Summary

This test was administered to 148 normal children with a mean age of 9.9 years and to a group of 49 emotionally ill children at a child guidance center. Mean age of the latter group was 10.0.

A teacher, principal and psychologist selected what appeared to be normal children, basing their judgment on a child's history, psychological tests and classroom behavior. The child guidance center cases were selected on the basis of age only. All of the child guidance center cases were judged to be emotionally ill by a psychiatrist, a psychologist and a social worker. The I. Q.'s of both groups were matched. A KTSA was individually administered. Average administration time was 15 minutes; average scoring time was 4 minutes.

The KTSA appears to be an effective and convenient instrument for the initial screening of emotionally ill and emotionally healthy children. In addition to the objective scoring criteria, there were many projections on the test which were helpful in understanding the nature of a child's difficulty.

REFERENCES

1. Kahn, T. C. "An Original Test of Symbol Arrangement Validated on Organic Psychotics," *Journal of Consulting Psychology*, 1951, 15, pp. 439-444.
2. Kahn, T. C. "Personality Projection on Culturally Structured Symbols," *Journal of Projective Techniques*, 1955, 19, pp. 431-442.
3. Kahn, T. C. "Kahn Test of Symbol Arrangement; Administration and Scoring," *Perceptual and Motor Skills*, 1956, 6, Monograph 1.
4. Kahn, T. C. "Kahn Test of Symbol Arrangement; Clinical Manual," *Perceptual and Motor Skills*, 1957, 7, Monograph 1.
5. Kahn, T. C. and Murphy, P. H. "A New Symbol Approach to Personality Assessment," *The American Journal of Psychiatry*, accepted for publication.

A STUDY OF PREDICTIVE VALIDITY OF THE MINNESOTA TEACHER ATTITUDE INVENTORY

HARRY P. DAY
Florida State University
Tallahassee, Florida

The purpose of this study was to determine the predictive validity of the MTAI¹ for forecasting teaching effectiveness after one year of in-service experience.

Although validity of a concurrent kind was determined by the test authors by comparing contemporary performance and contemporary MTAI performance, there is a lack of information about the predictive validity of the instrument for certain purposes. Such evidence is crucially important if one wishes to use the MTAI as a basis for selecting teacher training candidates.

Procedure

The MTAI was administered to 196 college seniors, 40 males and 156 females, immediately upon their return to campus from internship teaching service in the public schools. Of this group, 87 were elementary interns and 109 were secondary interns.

Then one year later a follow-up study was conducted. Of the 196 who had completed their internship, 135 were employed full-time as teachers.

The MTAI was mailed to all of the original group who could be located with the request that they administer the MTAI to themselves and return the booklet and answer sheet to the investigator. Of the 135 who were teaching, responses were received from 109 subjects. Of the 61 who were not teaching, 37 also returned the completed materials.

For each of the 109 subjects who were teaching, and had returned their completed forms, two ratings of their in-service teaching effectiveness were sought through the cooperation of each teacher's principal and supervisor. The principal's rating of the teacher's effectiveness was determined by having the principal complete the Principal-Teacher Rating Scale² for the teacher. The Supervisor's rating was determined by having the supervisor complete the Baxter Scale for Rating Teachers' Personal Effectiveness as modified by Leeds.³ For the 109 teachers who returned the MTAI retest, 70 complete sets of principal and supervisor ratings were collected.

Results

Complete information--post-internship MTAI, MTAI after one year of teaching, principal's rating of effectiveness, and supervisor's rating of effectiveness--was gathered for 70 of the 135 subjects who had taken teaching positions following their graduation from college.

Table I contains the intercorrelations of the certain variables on which information was collected from the teaching group.

TABLE I

INTERCORRELATIONS OF DATA FROM GROUP
WHO BECAME TEACHERS*

Test	MTAI (I)	Principal's Rating
MTAI (II)	.63	
Principal Rating	.28	
Supervisor Rating	.18	.62

*N=70

It appears from the data collected that the predictive validity of MTAI scores when collected immediately after internship is quite low when the criterion is principal's ratings of effectiveness and when the ratings are collected after one year of service or when the criterion is supervisor's ratings.

This does not, however, reflect discredit on the goodness of the instrument. The MTAI purports only to assess attitudes the teacher holds toward the pupils. The attitude which a teacher holds about proper teacher-pupil relationships is admittedly important, but it certainly cannot be expected to reveal the totality of a teacher's effectiveness or ineffectiveness. Teacher effectiveness is a complex of many factors. Possibly one might study the value of the MTAI as a

* Footnotes will be found at the end of the article.

predictive measure when checked against other criteria and when used with other measures.

FOOTNOTES

- * This article is based on a portion of the writer's doctoral dissertation: Day, Harry P. A Study of the Validity of the Minnesota Teacher Attitude Inventory as a Predictive Instrument in the Selection of Good Teaching Prospects from among College Undergraduates, Unpublished Ed. D. dissertation, School of Educa-

- tion, Florida State University, 1956, p. 117.
1. Cook, Walter W., Leeds, Carroll H., and Callis, Robert, Minnesota Teacher Attitude Inventory (New York: The Psychological Corporation, 1951).
 2. Leeds, Carroll H., The Construction and Differential Value of a Scale for Determining Teacher-Pupil Attitudes, Unpublished Doctor's dissertation, University of Minnesota, 1946. (See Appendix III).
 3. Leeds, Carroll H., "A Second Validity Study of the Minnesota Teacher Attitude Inventory," Elementary School Journal, LII (1952), p. 401.

BOOK REVIEWS

L. Joseph Lins, University of Wisconsin; Valworth Plum, University of Minnesota;
John Schmid, Michigan State College

Brief Commentaries

OTIS, ARTHUR S. Added Revenue without Burden (Boston: The Christopher Publishing House 1958), pp. 123, \$3.00.

Otis presents a plan of taxation based upon the premise that land values are rising and that the increased rental value of land is a source of revenue that may be drawn upon without overburdening the landowner. The plan is to have the tax based upon the assessed capital value.

L. J. L.

PEARSON, C. ERIC. A Classroom Teacher's Guide to Physical Education (New York: Bureau of Publications, Teachers College, Columbia University, 1958), pp. x+127.

A problem-type approach is used in the development of suggestions for the teaching of physical education. This is a very good treatment of physical education as preparation for physical activity throughout life.

L. J. L.

PRESTON, RALPH C. Teaching Social Studies in the Elementary School (New York: Rinehart and Company, 1959), pp. 382, \$5.00.

This is a revision of a book which first appeared eight years ago. More attention is given in this revision to social trends and values, research findings, pupil participation and committee work, creative experiences, and visual aids.

J. S.

ROBINSON, HELEN M. (Ed.) Materials for Reading (Chicago: University of Chicago Press, 1957), pp. vii + 231, \$3.50.

This is a composite of the reports at the Twentieth Annual Reading Conference which concentrated upon developing skills in reading and promoting personal and social growth through reading. Over forty-five reports delivered at the Conference, by persons from throughout the United States, are included.

L. J. L.

ROSENBERG, MORRIS. Occupations and Values (Glencoe, Illinois: The Free Press, 1957), pp. 158, \$4.00.

This book is a study of occupational choices of people in light of their individual and social values. Special emphasis is placed upon changes in occupational choices as a result of various factors; personality factors, degree of specialization, aspirations and expectations, career dedication, etc.

J. S.

SLOTE, CLAIRE TRIEB. Improve your Handwriting. (New York: McGraw-Hill Book Company, 1958), pp. x+147, \$3.50.

Stressing the importance of good handwriting, the author devotes sections of this book to methods of quickly and easily improving handwriting. A section traces the history of handwriting and the evolution of numbers from cavemen to current times.

L. J. L.

STRANG, RUTH. The Adolescent Views Himself (New York: McGraw-Hill Book Company, 1957), pp. 581, \$7.95.

This is a practical book which should enable secondary school teachers to better understand adolescence. This book contains many personal statements of adolescents which show their attitudes and values, their activities and relationships, and their problems of growing up. This reviewer endorses whole-heartedly this book as one to be read by teacher and prospective teacher.

J. S.

WALKER, ROBERT H. American Studies in the United States. A Survey of College Programs (Baton Rouge: Louisiana State University Press, 1958), pp. 210, \$3.00.

This is a report of the results of a questionnaire sent to institutions with activity in American Studies, of descriptive essays relative to the American Studies programs of the various institutions,

and of opinions expressed at a conference on American Studies held in Washington D. C.
L. J. L.

WILLEY, ROY DeVERL; STRONG, W. MELVIN
Group Procedures in Guidance (New York: Harper and Brothers, 1957), pp. 548, \$6.00.

This is a textbook for use by prospective and practicing school personnel. It treats group procedures. Part I deals with the place of guidance in education. Part II deals with group approaches to guidance. Part III treats group guidance in core course. A supplement is provided which serves as an example of an appropriate unit for a core

course: appreciating the contribution of other cultures.
J.S.

YUKER, HAROLD E. A Guide to Statistical Calculations (New York: G. P. Putnam's Sons, 1958), pp. viii + 95, \$1.95.

This manual is designed primarily as a help to the person without college mathematics. Elementary statistical formulas are presented with each statistic being presented in the form of a problem. The solution of the problem is outlined in a step sequence. The book deals with the calculation of a statistic not with when a particular statistic should be used or the logic underlying its use.
L. J. L.

JOURNAL *of* EDUCATIONAL RESEARCH

VOLUME 53

OCTOBER, 1959

NUMBER 2

CONTENTS

A Comparative Analysis of Personality and Ego Strength Test Scores for In-Prison, Neuro-Psychiatric and Typ- ical Individuals—Russell N. Cassel	43
Stimulating Interest in Public Health Problems Among High School Pupils—Helen L. Williams, Warren H. Southworth.....	53
Rural-Urban Differences in Intelligence—Irvin J. Lehmann.....	62
College Students' Ability to Evaluate Their Performance on Objective Tests—Earl W. Kooker, Chester S. Williams	69
Measurement of Teacher Merit for Salary Purposes— William A. McCall, Gertrude R. Krause.....	73
A Study of the Effectiveness of a Summer Remedial Course in English for College Freshmen—Martin L. Zeigler, Louis M. Herman	76
An Analysis of Test Scores and Grades for Predicting Suc- cess of College Students in English Composition—Wil- liam E. Kunhart, Lionel R. Olsen.....	79

DEMBAR PUBLICATIONS, INC.

Madison 3, Wisconsin

A COMPARATIVE ANALYSIS OF PERSONALITY AND EGO STRENGTH TEST SCORES FOR IN-PRISON, NEURO-PSYCHIATRIC AND TYPICAL INDIVIDUALS

RUSSELL N. CASSEL, Fontana, California
B. LYNN HARRIMAN, Federal Correctional Institution, Englewood, Colorado

THIS STUDY IS concerned with making a comparative analysis of scores from a personality tension and needs test, namely, the Group Personality Projective Test; and an ego-strength test, namely, the Ego-Strength Q-Sort Test, for in-prison, neuro-psychiatric and typical subjects. It is a portion of a larger study directed at the investigation of psychological dynamics present in the delinquency-prone youth.

The larger study involves the efforts of four separate groups of task scientists working in centers in New York City; Cincinnati, Ohio; Oklahoma and Southern Texas; and Phoenix, Arizona. Four separate areas of investigation with expressed objectives as follows are included: 1) home and family impoverishment-degree to which home contributes to delinquency, and level of adjustment essential for return to home; 2) personality need and ego-strength-degree of mental health and self-potential for dealing effectively with own problems; 3) social insight and irreality assessment-characteristic reactions of individuals to a wide range of social problem situations, and degree of irreality of perceptions as an index of social sensitivity; and 4) leadership patterns and potential-determine means for optimum development of the individual.

Progress reports have been made at recent annual meetings of both the American Psychological Association and the American Correctional Congress on varying aspects of the larger study. At the present time, psychological tests have been developed and standardized for each of the four referenced areas, and for the most part these tests are published by recognized companies and available for purchase and use. To date, no effort has been made to utilize the entire battery in the ultimate intended sense, but plans are under way for such investigation in the future.

Tests Utilized

Personality Tension and Needs: The Group Personality Projective Test is currently published and distributed exclusively by Psychological Test Specialists, Box 1441, Missoula, Montana. The two

principal features of the test are 1) projective structure, and 2) factorial validity. A third important, but not a principal, feature is the assessment of percentage of negative feelings projected.

Ninety stick-figure drawings are used as a means of soliciting responses relative to a widely diversified range of usual life activities. Each of the drawings has a minimum of meaning or paucity of structuring which is designed to provide an opportunity for the subject to utilize active need areas in their own personality to temper or complete the structuring, or as a basis to select an answer. Thus, through this paucity of structuring, the subject reveals those areas of active needs utilized in selecting an answer and thereby contributes to the assessment of the most cogent and active personality needs for the moment of time.

Six part scores and a total score are provided which are defined as follows: 1) tension-percentage of negative feelings projected, 2) nurturance-need to play father role or to give aid to others, 3) withdrawal-need to escape or desire for non-participation in activity, 4) affiliation and psychosexual-need to belong and for psychosexual (boy-girl) relationships, and 5) succorance-need to play mother role or to seek aid. Total score-total personality needs or disturbance.

Ego Strength: The Ego Strength Q-Sort Test is a test published by Psychometric Affiliates, 221 North La Salle Street, Chicago 1, Illinois. It is designed primarily to determine the nature and degree to which a person's "ego-ideal" is developed, the relative importance of certain factors pertinent to this ideal, and the present state of ego-strength of the individual.

For purposes of this test, the ego is defined as that aspect of the personality which comes into contact with reality, and which has the chief function of generating and maintaining happiness and contentment for the person. No differentiation is made between an ego which has been seriously wounded in onslaught with the environment and one that is inadequate or retarded in development.

The primary feature of the test is that it employs Q-methodology as a means to interpret the main

score, and R-methodology as a means to determine the test profile (part scores), or reason for the main score. The five part scores of the test are profiled on the basis of T-score values which were obtained by summing the average ratings assigned to the related items for each of the separate groups or clusters (five part scores), and by comparing those ratings with similar mean ratings for widely diversified groups.

The five part scores are defined as follows: 1) ego status-self sufficiency, 2) social status-personal prestige, 3) goal setting and striving-ego involvement, 4) good mental health-personal adjustment, and 5) physical status-good physical health.

The total score is obtained by correlating the Q-sort array (total ratings assigned to the 60 Q-sort items in terms of a forced choice normalized pattern) made by the subject taking the test with the mean ratings assigned by widely diversified groups of individuals (test norm); all in terms of ratings indicating the persons opinion of the value which each of the 60 Q-sort items have in generating or maintaining happiness in the individual. This score is presumed to be an indication of the present potential of an individual to deal with his own problems.

Groups Utilized

Typical Personality Groups

Select Youth (male): One hundred individuals ranging in age from 17 to 26 years old, and with a M age of 17.32 and a SD of 3.79 years. In terms of scholastic achievement, the last school grade successfully completed ranged from 12 to 19, and with a M grade of 13.20 and a SD of 0.92 grades. Their IQ's ranged from 100 to 150 with a M of 117.30 and with a SD of 8.91. All of these individuals were Air Force Pre-flight Cadets and they were carefully screened on the basis of high intellectual, educational, and physical qualifications.

Unselected Adults (mixed sex): One hundred young and middle-aged adults, about equally divided for sex, ranging in age from 14 to 50 years old, and with a M age of 33.71 and a SD of 13.3 years. Last school grade successfully completed ranged from 8 to 15, and with a M grade of 9.8 and a SD of 1.7 grades. Their IQ's were in the average range. Specific scores were not available for all subjects. These individuals consisted of persons available to graduate students in the San Antonio, Texas area.

Ordained Religious Ministers: Thirty young male adults about half of whom were Catholic priests, and the remainder came from a wide diversity of Protestant religions; two of whom were Jewish Rabbis. All of these individuals had a minimum of five years of college training, the completion of their respective religious ministerial training, and about five years of experience in the ministry of their respective faiths. Their average age was from 25 to 35 years, and in all cases their IQ's were upward of 115. No further identification

is permissible, for obvious professional reasons.

Deviant Personality Groups

Neuro-Psychiatric Hospital Patients: One hundred patients in two general hospitals, both with inpatient populations in excess of 3,000, and both situated in southwestern United States. All of these patients were diagnosed as having schizophrenia-type maladies as a basis for hospitalization (mostly simple stage and ambulatory), and about equally divided between the sexes. They ranged in age from 14 to 60 years with a M of 37 and a SD of 4.7 years. All of them had a minimum of 5th grade education or better. They were taken from a hospital roster of names, eliminating those who were not classified as schizophrenia, or those who were unable to complete the tests. In general, they were of average intelligence, but no specific scores were available relative to IQ's.

Federal Reformatory Prisoners (male): One hundred male adult prisoners currently committed to a Federal Reformatory in central United States, and all of whom volunteered to take experimental editions of personality and leadership tests. They ranged in age from 17 to 27 years with a M age of 22.3 and with a SD of 3.7 years. Their last school grade completed ranged from 6 to 16 with a M grade of 10.00 and a SD of 2.7. Their M IQ was 92 with a SD of 14.2.

Findings

Personality Tension and Needs: Comparisons for this portion of the study were based on the difference between means contained in Table I, and only significant differences are noted.

Selected Youth vs. Unselected Adults:

1. Less tension in Selected Youth (t 18.3).
2. Less withdrawal in Selected Youth (t 12.03).
3. Less neuroticism in Selected Youth (t 9.02).
4. More affiliation in Selected Youth (t 12.82).
5. Less succorance in Selected Youth (t 4.20).
6. Less total disturbance in Selected Youth (t 20.12).

Selected Youth vs. Neuro-psychiatric Patients:

1. Less tension in Selected Youth (t 26.42).
2. Less nurturance in Selected Youth (t 9.96).
3. Less withdrawal in Selected Youth (t 9.04).
4. Less neuroticism in Selected Youth (t 2.91).
5. More affiliation in Selected Youth (t 10.50).
6. Less succorance in Selected Youth (t 2.34).
7. Less total disturbance in Selected Youth (t 12.50).

Selected Youth vs. Federal Reformatory Prisoners:

1. Less tension in Selected Youth (t 15.94).
2. Less withdrawal in Selected Youth (t 10.43).
3. Less neuroticism in Selected Youth (t 3.84).
4. More affiliation in Selected Youth (t 14.81).
5. Less succorance in Selected Youth (t 19.73).
6. Less total disturbance in Selected Youth (t 12.50).

Selected Youth vs. Ordained Religious Ministers:

1. Less tension in Selected Youth (t 3.38).
2. Less withdrawal in Selected Youth (t 4.62).

TABLE I
COMPARATIVE PART AND TOTAL SCORES ON THE GROUP PERSONALITY PROJECTIVE TEST FOR
BOTH THE TYPICAL AND DEVIANT PERSONALITY GROUPS OF INDIVIDUALS

Groups	Tension Reduction Quotient	Nurturance Score	Withdrawal Score	Neuroticism Score	Affiliation Psycho-Sexual Score	Succorance Score	Total Personality
<u>Typical Personality Groups</u>							
Select Youth (male)(N-100)							
Mean	21.10	8.8	9.7	17.40	21.5	8.42	45.9
SD	11.00	2.5	2.5	4.3	4.4	3.70	10.6
Unselected Adults (mixed sex) (N-100)							
Mean	39.17	8.89	13.07	21.55	16.47	9.89	62.01
SD	13.24	2.47	3.75	6.25	5.12	4.75	9.88
Ordained Religious Ministers (N-30)							
Mean	28.40	9.64	12.24	18.75	17.03	10.93	53.40
SD	11.50	2.62	2.90	3.72	3.24	2.17	10.41
<u>Deviant Personality Groups</u>							
Neuro-Psychiatric Hospital Patients (mixed sex)(N-100)							
Mean	46.20	11.49	11.87	23.80	17.30	9.28	68.60
SD	12.18	3.49	3.29	6.18	6.09	5.05	10.30
Federal Reformatory Prisoners (N-100)							
Mean	36.80	8.3	12.1	18.4	17.5	11.2	61.40
SD	17.30	3.5	3.4	3.4	3.6	4.3	19.50

TABLE II
PREDICTION VALIDITY (BY USE OF CROSS VALIDATION b WEIGHTS) OF
WEIGHTED COMPOSITE SCORE ON GPPT FOR DISCERNING BETWEEN
TYPICAL AND DEVIANT PERSONALITY GROUPS

Prediction Composite Score	Typical Individuals		In-Prison Subjects			N-P Patients		
	No.	Cum.	No.	Cum.	Prediction Accuracy	No.	Cum.	Prediction Accuracy
150-159			4	4				
140-149			2	6				
130-139			0	6		1	1	
120-129			3	9		3	4	
110-119			2	11		2	6	
100-109			2	13		5	11	
90- 99			5	18		27	38	
80- 89	2	558	20	38	72%	50	88	82%
70- 79	8	556	26	64	74%	93	181	93%
60- 69	64	548	84	148	76%	35	216	89%
50- 59	136	484	55	203	66%	10	226	73%
40- 49	216	348	50	253	46%	2	228	46%
30- 39	96	132	15	268		1	229	
20- 29	32	36	4	272		1	230	
10- 19	4	4						

Note: accuracy figures are read below the line for typical individuals and above the line for in-prison and N-P patients. The use of twice as many (approximately) typical subjects as for the deviant personality groups is believed to be somewhat representative of the characteristic populations for which psychological tests are usually used.

TABLE III

STATUS VALIDITY POINT BI-SERIAL CORRELATION COEFFICIENTS FOR PART AND TOTAL SCORES ON THE EGO
STRENGTH Q-SORT TEST FOR DISCERNING BETWEEN TYPICAL AND DEVIANT PERSONALITY GROUPS

Scores	Typical vs. N-P Patients		Typical vs. Federal Prisoners		Typical vs. Religious Ministers		Federal Prisoners vs. N-P Patients	Federal Prisoners vs. Religious Ministers	N-P Patients vs. Religious Ministers	Unselected Adults vs. Selected Youth
	Unselected Adults	Selected Youth	Unselected Adults	Selected Youth	Unselected Adults	Selected Youth				
<u>Part Scores</u>										
Ego-Status	.091	-.217**	.151*	-.224*	-.122	.328**	-.038	.137	.081	-.345**
Social Status	-.223**	.276**	-.094	.204**	.408**	-.073	.088	.24	.200*	.484**
Goal Setting	.250**	.018	.337**	.106	-.271**	-.062	-.101	.040	-.051	-.200**
Mental Health-	.111	-.235**	-.065	-.196**	-.172*	-.007	-.040	-.205*	-.245**	-.160*
Physical Status	.049	-.158	.149*	.108	-.238**	-.193*	-.105	-.079	-.015	-.046
<u>Total Scores</u>										
# R- Methodology	.293**	.404**	.341**	.341	.406**	.326**	.076	.342**	.255**	.782**
Q- Methodology	.210**	.605	.017	.415**	.555**	.102	.365**	.072**	.683**	.440*
# Multiple R_{pb}										
* Significant at the 5% level										
**Significant at the 1% level										

3. More affiliation in Selected Youth (t 7.26).
4. Less succorance in Selected Youth (t 5.95).
5. Less total disturbance in Selected Youth (t 5.95).

Federal Prisoners vs. N-P Patients:

1. Less tension in Federal Prisoners (t 6.96).
2. Less nurturance in Federal Prisoners (t 13.5).
3. Less succorance in Federal Prisoners (t 4.36).
4. Less total disturbance in Federal Prisoners (t 3.54).

Religious Ministers vs. N-P Patients:

1. Less tension in Religious Ministers (t 7.80).
2. Less nurturance in Religious Ministers (t 3.36).
3. Less neuroticism in Religious Ministers (t 6.16).
4. More succorance in Religious Ministers (t 3.11).
5. Less total disturbance in Religious Ministers (t 7.38).

Ego Strength: Comparisons for this portion of the study were based on inter-comparisons between the point bi-serial correlation coefficients from Table III and the means from Table IV, and using only indices which were statistically significant at the 5% level of confidence or better.

Selected Youth vs. Unselected Adults:

1. More emphasis on self and ego status for Selected Youth.
2. Less emphasis on social status for Selected Youth.
3. More emphasis on goal setting and striving for Selected Youth.
4. Less emphasis on mental health and adjustment for Selected Youth.

Selected Youth vs. N-P Patients:

1. More emphasis on self and ego-status for Selected Youth.
2. Less emphasis on social status for Selected Youth.
3. Less emphasis on mental health for Selected Youth.
4. More emphasis on physical status for Selected Youth.

Selected Youth vs. Federal Reformatory Prisoners:

1. More emphasis on self and ego-status for Selected Youth.
2. Less emphasis on social status for Selected Youth.
3. Less emphasis on mental health for Selected Youth.

Selected Youth vs. Ordained Religious Ministers:

1. More emphasis on self and ego-status for Selected Youth.
2. Less emphasis on Physical status for Selected Youth.

Unselected Adults vs. N-P Patients:

1. More emphasis on social status for Unselected Adults.
2. Less emphasis on goal setting for Unselected

Adults.

Unselected Adults vs. Federal Prisoners:

1. Less emphasis on goal setting and striving for Unselected Adults.
2. Less emphasis on physical status for Unselected Adults.

Unselected Adults vs. Ordained Religious Ministers:

1. More emphasis on social status for Unselected Adults.
2. Less emphasis on goal setting and striving for Unselected Adults.
3. Less emphasis on physical status for Unselected Adults.

Federal Prisoners vs. N-P Patients:

1. No significant differences for any of the five part scores.

Federal Prisoners vs. Ordained Religious Ministers:

1. Less emphasis on mental health and adjustment for Federal Prisoners.

N-P Patients vs. Ordained Religious Ministers:

1. More emphasis on social status for N-P Patients.
2. Less emphasis on physical status for N-P Patients.

For generalized findings, see next page.

Conclusions

Personality Tension and Needs:

1. Tension (as measured by percentage of negative projections) is pertinent in discerning personality variables relative to Selected Youth and N-P Patients.
2. Nurturance is pertinent in discerning relative to N-P Patients and Religious Ministers.
3. Withdrawal is pertinent in discerning relative to Selected Youth.
4. Neuroticism is pertinent in discerning relative to Selected Youth and N-P Patients.
5. Affiliation is pertinent in discerning only relative to Selected Youth.
6. Succorance is pertinent in discerning relative to Selected Youth, Federal Prisoners, and Religious Ministers.
7. Total personality disturbance may be used reliably (with both statistical and practical significance) for discerning between typical and deviant personality patterns, but has little practical value for discerning groups that may be independently organized within the typical or deviant classification structure, i.e., one can discern the N-P or potential delinquent from the typical individual but not the N-P from the potential delinquent.

Ego-Strength:

1. Ego-status ratings are pertinent for discerning the ego-strength patterns relative to Selected Youth, N-P Patients, Federal Prisoners, and Religious Ministers.
2. Social status ratings are pertinent relative to Selected Youth and N-P Patients.

GENERALIZED FINDINGSPersonality Tension and NeedsSelected Youth

1. Low tension
2. Low withdrawal
3. Low neuroticism
4. High affiliation
5. Low succorance
6. Low total disturbance

N-P Patients

1. High tension
2. High nurturance
3. High neuroticism
4. High total disturbance

Federal Prisoners

1. Moderate tension
2. High succorance
3. High total disturbance

Religious Ministers

1. High nurturance
2. High succorance
3. Low total disturbance

Ego-StrengthSelected Youth

1. High ego-status
2. Low social status
3. Low mental health
4. High ego-strength

N-P Patients

1. High social status
2. Moderate mental health
3. Extremely low ego-strength

Federal Prisoners

1. High mental health
2. Moderate ego-strength

Religious Ministers

1. High mental health
2. High ego-strength

TABLE IV
COMPARATIVE PART AND TOTAL SCORES ON THE EGO STRENGTH Q-SORT TEST
FOR BOTH THE TYPICAL AND DEVIANT GROUPS OF INDIVIDUALS

Groups	Part Scores					Total Score
	Ego- Status	Social Status	Goal Setting	Mental Health	Physical Status	Ego- Strength (Fisher z')
<u>Typical Personality Groups</u>						
Selected Youth (Male)(N-100)						
Mean	61.8	88.2	80.8	111.3	13.5	.473
SD	4.5	8.6	11.5	8.7	3.6	.155
Unselected Adults (Mixed sex) (N-100)						
Mean	57.9	97.5	76.4	114.9	12.3	.260
SD	6.0	7.9	9.3	5.7	3.6	.180
Ordained Religious Ministers (N-30)						
Mean	58.1	89.5	82.1	117.4	14.1	.499
SD	6.3	9.5	10.6	8.3	3.5	.100
<u>Deviant Personality Groups</u>						
Neuro-Psychiatric Hospital Patients (Mixed sex)(N-100)						
Mean	59.1	93.4	81.1	113.2	12.7	.189
SD	6.9	9.3	8.8	8.5	3.6	.132
Federal Reformatory Prisoners (N-100)						
Mean	59.6	91.8	82.9	117.3	12.7	.298
SD	5.0	8.6	8.7	8.7	3.8	.152

TABLE V

PREDICTION VALIDITY (BY USE OF CROSS VALIDATION b WEIGHTS) OF
WEIGHTED COMPOSITE SCORE ON ESQT FOR DISCERNING BETWEEN
TYPICAL AND DEVIANT PERSONALITY GROUPS

Fisher z' Scores	Typical Individuals		Federal Prisoners		N-P Patients		N-P Patient Prediction		
	No.	Cum.	No.	Cum.	No.	Cum.	Two - way Accuracy	%age of N-P Accurate	%age of Error
75 to 79	6	6							
70 to 74	6	12	5	225					
65 to 69	3	15	5	220					
60 to 64	24	39	10	215					
55 to 59	24	63	5	205	7	211			
50 to 54	33	96	15	200	8	204			
45 to 49	18	114	5	185	0	196			
40 to 44	18	132	15	180	0	196	77%	93	33
35 to 39	12	144	25	165	11	196	77%	83	33
30 to 34	15	159	10	140	11	185	80%	82 "Cut-off"	20
25 to 29	15	174	15	130	33	174	74%	67	26
20 to 24	21	195	30	115	33	141	71%	51	29
15 to 19	3	198	10	85	17	108			
10 to 14	8	207	35	75	33	91			
05 to 09	6	213	35	40	22	53			
00 to 04	3	216	0	5	17	36			
00 to -04			5	5	0	19			
-05 to -09					11	19			
-10 to -14					8	8			
-15 to -19									

Above "Cut-off"

Relatively Strong Egos

Below "Cut-off"

Relatively Weak Egos

3. Goal setting ratings are not pertinent relative to Selected Youth, N-P Patients, or Religious Ministers.

4. Mental Health ratings are pertinent relative to Selected Youth, Federal Prisoners, and Religious Ministers.

5. Ego-strength may be utilized effectively for discerning between the typical individuals and N-P patients, but has little practical value in discerning the potential delinquent.

Summary

This study is concerned with making a comparative analysis of scores from the GPPT and the ESQT for N-P Patients, Federal Reformatory Prisoners, and varying groups of typical subjects. It is a portion of a larger study dealing with the prediction and control of delinquency proneness in youth. The larger study involves four separate areas of investigation: 1) home and family impoverishment, 2) personality needs and ego strength, 3) social insight and irreality assessment, and 4) leadership patterns and potential.

In terms of personality tension and needs, tension appears to be pertinent to Selected Youth and N-P Patients; nurturance appears to be pertinent to N-P Patients and Religious Ministers; withdrawal is pertinent to Selected Youth; affiliation is pertinent to Selected Youth; succorance is pertinent to Selected Youth, Federal Prisoners, and Religious Ministers; and total personality disturbance is pertinent to all five groups of the study.

In terms of ego-strength, ego-status ratings appear to be pertinent to all five groups; social status to Selected Youth and N-P Patients; mental health to Selected Youth, Federal Prisoners, and Religious Ministers; and ego strength to all five groups.

The GPPT and the ESQT may be utilized for discerning with both statistical and practical significance between individuals and groups of typical and deviant personality patterns; but with little practical significance between N-P Patients and Federal Prisoners. Both the personality and ego-strength scores for the N-P Patients are more homogenous than are similar scores for Federal Prisoners. In general, the personality scores show more serious disturbance, and the ego-strength lower ego-strength for the N-P Patients than for the Federal Prisoners.

REFERENCES

1. Cassel, R. N., and Kahn, T. C., The Group Personality Projective Test and Manual, Psychological Test Specialists, Box 1441, Missoula, Montana, 1957.
2. Cassel, R. N., The Ego Strength Q-Sort Test and Manual, Psychometric Affiliates, 221 North LaSalle Street, Chicago, Illinois, 1958.
3. Cassel, R. N., and Kahn, R. C., "Development and Standardization of the Group Personality Projective Test", Journal of Projective Techniques, Vol. 22, No. 3, 1958.
4. Cassel, R. N., "Comparing the Effectiveness of the Ego Strength Q-Sort Test by Use of R- and Q-methodologies", Journal of Genetic Psychology (pending).
5. Hartman, B., Developing a Functional Definition of Ego Strength by Factor Analyzing Available Tests, Unpublished Master's Thesis, Our Lady of the Lake College, San Antonio, Texas, 1957.
6. Horst, P., and Smith, S., "The Discrimination of Two Racial Samples", Psychometrika, Vol. 15, No. 3, September, 1950.

STIMULATING INTEREST IN PUBLIC HEALTH PROBLEMS AMONG HIGH SCHOOL PUPILS

HELEN L. WILLIAMS, Memorial High School, Beloit, Wisconsin
WARREN H. SOUTHWORTH, University of Wisconsin

PUBLISHED STUDIES of the health interests of high school pupils are not numerous. Those that have been made by Warren H. Southworth in Massachusetts (6), by Joseph Lantagne in selected schools sampling various regions of the entire United States (3), and by the Public Schools in Denver, Colorado (2), indicated that, for this age group, interest in community and world health problems is decidedly lower than interest in personal health. While this self-centeredness may be natural in teen-agers, the authors believed that an interest in public health matters could be increased by specific teaching. To determine to what extent such teaching would be successful, an experiment was undertaken in four sections (classes) of biology pupils in the Beloit high school during the academic year of 1956-7.

Beloit is a city of approximately 32,000 population in the heart of the mid-west. Its principal industries are the manufacture of machine tools, motors, shoes, and the processing of dairy and other food products. It is the site of a small college, and is surrounded by a rich agricultural area. There is a variety of ethnic groups, including New Englanders, Norwegians, Irish, Italians, and Negroes. Approximately 115 pupils, mostly tenth graders, participated in the study. One section contained pupils with IQ's of 115 and over; two sections were made up of pupils with 90 to 115 IQ; and one section was of pupils with 80 to 90 IQ.

The Experiment

The test-retest method seemed most appropriate for judging the validity of the hypothesis (that interest in public health could be increased by specific teaching). An interest inventory of 125 items was prepared, with about two thirds of the items relating to personal health and one third to public health, although the items were not so identified to the pupils. The following sources were used:

The Massachusetts study by Southworth, Latimer, and Turner (6).

A summary of the Denver study (2).

Mooney Problem Check-list (5).

Moon, Mann, Otto--Modern Biology (text used in the course) (4).

This inventory was first administered to the pupils in April, 1957. An introductory paragraph read: "Any subject is more valuable to those studying it if it deals principally with topics that the students are interested in. This list of questions is to try to find out what you and your classmates are interested in. Mark an X in front of those questions that you would like to learn the answers to. Do not answer the questions now. Do not sign your name."

Difficult words were pronounced for the pupils if they requested this help, but the words were not defined and no other help or explanations were given. Most pupils took 20 to 30 minutes to mark their choices.

The classes then undertook the study of the text and related materials on health, of which the following is a brief outline:

I. Causes of Disease, Disability, and Death.

A. Infectious Diseases.

1. Virus Diseases.
2. Rickettsial Diseases.
3. Bacterial Diseases.
4. Spirochaetal Diseases.
5. Protozoan Diseases.
6. Fungus Diseases.
7. Diseases caused by insects and related organisms.
8. Diseases caused by worms and related organisms.

For each type of infection, the following sub-points were covered:

1. Nature of the organism.
2. Method of spread.
3. Effect on the body.
4. Methods of control.

B. Types of poisoning--metals, alkaloids, allergens.

C. Other External Causes (environmental).

1. Variations in Atmospheric Temperature.
2. Variations in Atmospheric Pressure.
3. Sound and Ultrasonic Vibrations.
4. Light.

- 5. Radiations.
- D. Chronic and Degenerative Diseases.
- E. Hormone Imbalance.
- F. Hereditary Diseases.
- G. Congenital Diseases.
- H. Mental and Emotional Disorders.
- I. Deficiency Diseases (especially mineral and vitamin deficiencies).
- II. Treatment of Diseases.
 - A. Natural Body Defenses: nature and use of each.
 - 1. Skin and Mucous Membranes.
 - 2. Blood Cells and Antibodies.
 - 3. Fever.
 - 4. General Resistance.
 - B. Immune Therapy.
 - 1. Nature of Immunity; How Acquired.
 - 2. Production of Immunizing Materials.
 - 3. Importance of Community-Wide Immunization Programs.
 - C. Specifics.
 - 1. Antibiotics.
 - 2. Chemotherapy.
 - 3. Hormones.
 - 4. Vitamins.
 - D. Rest and Relaxation.
 - E. Surgery.
 - F. Exercise and Physical Therapy.
 - G. Psychiatry, Occupational Therapy, Shock, and Hydrotherapy (mostly for mental illness).
- III. Prevention of Disease.
 - A. Individual Health Behavior.
 - B. Environmental Sanitation (a community responsibility).
 - C. Pure Food and Drug Laws.
 - D. Medical Examinations (individual responsibility).
 - E. Proper Mental and Emotional Attitude (individual responsibility).
 - F. Temperance, especially in the use of alcohol and drugs (individual responsibility).
 - G. Safety in Driving, Work, and Play (both individual and community).
 - H. Immunization Programs (both).
 - I. Quarantine Regulations (both).
 - J. Education and Research (both).
 - K. Control of Radiations (community).

At all times during the teaching of this unit, in class discussion and in written work, reference was made to those activities which are individual responsibilities and to those which are community responsibilities. No unusual teaching methods were employed; three educational films on treatment and prevention of disease were shown, and there were some individual reports given on a voluntary basis. It would appear that to awaken interest in community health problems, it is only necessary to make pupils aware that such problems exist.

Many of the pupils noticed and commented that there are only a few activities that are strictly private or strictly public responsibilities; if the community is responsible for providing a health service

(e.g., immunization programs), then the individual is responsible for taking advantage of that service; and if the individual is responsible for certain activities (e.g., maintaining personal cleanliness), then the community is responsible for making it convenient for him to perform those activities (by providing an adequate water supply). More and more they came to realize that "no man is an island."

At the conclusion of the teaching and learning activities, the interest inventory was presented again; the introductory paragraph now read: "It is a good idea to look back over our work from time to time to see if what we are learning is important and useful. Here is a chance to check up on some of the things you have learned. Mark an X in front of those questions that you would like to learn more about. Do not answer the questions. Do not sign your name."

A summary of the results of the two administrations of the inventory is given in Table I. The first column of figures represents the percentage of the total group of pupils who marked each item on the first administration; the second column shows the percentage of pupils who marked it on the second administration; and the third column shows the percent of gain or loss in interest, derived by using the formula $\frac{A_2 - A_1}{A_1}$.

The results can be summarized as follows:

COMMUNITY HEALTH ITEMS. (N = 43)

Number showing gain in interest.....	31	(72.1%)
Number showing loss in interest.....	12	(27.9%)
Greatest Gain (Item 2)	428.0%	
Least Gain (Item 75)	0.5%	
Average Gain	111.9%	
Greatest Loss (Item 35).....	39.7%	
Least Loss (Item 60).....	4.0%	
Average Loss	22.9%	

PERSONAL HEALTH ITEMS. (N = 82)

Number showing gain in interest.....	47	(56.6%)
Number showing loss in interest.....	35	(43.4%)
Greatest Gain (Item 80).....	215.9%	
Least Gain (Item 4).....	1.9%	
Average Gain.....	46.5%	
Greatest Loss (Item 64)	46.1%	
Least Loss (Item 84).....	0.6%	
Average Loss.....	21.9%	

ITEMS OF GREATEST AND LEAST INTEREST

Personal item of greatest interest, first time, and 79--63.1%.	74
Personal item of greatest interest, second time, --52.8%.	30
Community item of greatest interest, first time, --56.1%.	12
Community item of greatest interest, second time,	73

TABLE I
SUMMARY OF THE RESULTS OF THE TWO ADMINISTRATIONS

Item No.	Personal or Community	Question	%, 1st Time	%, 2nd Time	% Gain or Loss
1	P	How can I find out how much I should weigh?	41.2	30.5	-25.9
2	C	Why must milk be pasteurized?	3.5	18.5	428.0
3	P	How can I keep from getting sick?	26.3	29.6	1.3
4	P	Why are some people "nervous"?	48.2	49.1	1.9
5	C	Why do some communities have health departments?	6.1	17.1	180.0
6	P	Are colds dangerous?	21.0	27.7	31.9
7	P	What causes pimples?	29.8	30.5	2.3
8	P	What is an allergy?	28.0	33.3	19.0
9	C	What responsibility does an individual have for the health of the community in which he lives?	17.5	27.7	58.3
10	C	What are the best ways of disposing of community wastes?	12.2	26.8	119.7
11	P	How can I care for my teeth?	18.4	22.2	20.7
12	P	What are venereal diseases?	56.1	39.8	-29.1
13	C	How do cities purify their water supplies?	14.9	23.1	55.7
14	P	How can I diet safely?	28.9	30.5	5.5
15	P	Is suntan healthful?	38.6	41.7	8.0
16	C	Do smoke and smog affect health or are they only nuisances?	46.5	38.1	-18.0
17	C	How do rats and mosquitoes spread disease?	10.5	19.4	84.8
18	P	Are frequent medical examinations really necessary?	14.9	20.4	36.9
19	P	Are patent medicines safe to use?	23.7	34.3	44.7
20	P	Is it really possible to "worry yourself sick"?	42.1	41.7	-1.0
21	C	What are some health hazards in industry and how can they be avoided?	9.6	22.2	131.2
22	C	How and why should flies and mosquitoes be destroyed?	5.3	16.6	213.2
23	P	Are all diseases caused by germs?	16.6	20.4	22.9
24	C	Why are births and deaths recorded by the state?	41.2	35.2	-14.6
25	P	What diseases are "incurable"?	41.2	31.5	-23.5
26	P	What diseases are "catching"?	31.6	26.8	-15.2

TABLE I (continued)

SUMMARY OF THE RESULTS OF THE TWO ADMINISTRATIONS

Item No.	Personal or Community	Question	%, 1st Time	%, 2nd Time	% Gain or Loss
27	C	What are the most destructive diseases in the world?	43.8	32.4	-26.0
28	C	Should employers require health examinations for all their workers?	7.9	20.4	158.2
29	P	What medicines can a person safely take by himself?	21.9	16.6	-24.2
30	P	Will cola beverages "rot your stomach"?	61.4	52.8	-14.0
31	C	What should be considered in selecting a family doctor or dentist?	22.8	25.9	13.6
32	P	Should everyone take vitamin pills?	20.2	32.4	60.4
33	P	Why should everyone have periodic chest x-rays?	7.9	17.6	122.8
34	P	How should one care for small cuts, burns, and bruises?	11.4	19.4	70.2
35	C	Why are blood tests required before marriage?	55.2	33.3	-39.7
36	C	What job openings are there in health work?	14.9	22.2	49.0
37	P	How do we inherit characteristics from our parents?	28.0	17.6	-37.1
38	P	Do cosmetics harm the skin?	43.8	36.1	-17.6
39	C	Should cities (governments) build hospitals or should this be left to private organizations?	8.7	27.7	218.4
40	C	What is geriatrics?	47.4	34.3	-32.4
41	P	How can athlete's foot be prevented?	15.8	20.4	29.1
42	P	What is the danger of lack of vitamins?	7.9	24.1	205.1
43	P	Is it harmful to be a blood donor?	15.8	20.4	29.1
44	C	How does the level of health in the United States compare with the level in other countries of the world?	16.6	27.7	66.9
45	P	How can one prevent heart disease?	31.6	30.5	-3.5
46	P	Does strenuous exercise harm the heart?	25.4	26.8	5.5
47	P	What causes food poisoning?	34.2	25.0	-26.9
48	P	Will "sun lamps" build up resistance to colds?	35.1	38.1	8.5
49	P	Does smoking affect one's health?	48.2	40.8	-15.4
50	C	Should schools require health examinations of all pupils?	11.4	23.1	102.6
51	P	What is a balanced diet?	7.9	13.9	75.9
52	C	What is the reason for the Pure Food and Drug Law?	28.9	25.9	-10.4

TABLE I (continued)

SUMMARY OF THE RESULTS OF THE TWO ADMINISTRATIONS

Item	Personal or Community	Question	%, 1st Time	%, 2nd Time	% Gain or Loss
53	P	Does massage help one to reduce?	26.3	34.3	30.4
54	P	What causes diabetes?	20.2	22.2	9.9
55	C	Should the government control the preparation of vaccines?	10.5	18.5	76.2
56	P	What causes fainting?	42.1	38.1	-9.5
57	P	What is the proper temperature to keep a study room?	12.2	20.4	67.2
58	P	What causes indigestion?	15.8	17.6	11.4
59	C	Why do cities have plumbing and electrical codes?	9.6	21.3	121.9
60	C	Who should take care of ill and aged people?	20.2	19.4	-4.0
61	P	Is it safe to use deodorants?	32.5	29.6	-8.9
62	P	How do we see and hear?	7.0	15.7	124.3
63	C	What are the main causes of death in the U.S.?	29.8	21.3	-28.5
64	P	How can one develop a pleasing personality?	39.5	21.3	-46.1
65	P	Does exercise build muscles?	26.3	23.1	-12.2
66	P	Is it harmful to go without breakfast?	39.5	24.1	-39.0
67	C	How has the Public Health Service reduced the death rate in the United States?	10.5	12.9	22.9
68	P	How do glands affect the body?	8.6	16.6	90.8
69	C	Is it cheaper to cure disease or to prevent it?	12.2	22.2	82.0
70	P	Is perspiration healthful?	13.1	24.1	84.0
71	P	Is fish a "brain food"?	26.3	21.3	-19.0
72	P	Are carrots good for the eyes?	25.4	21.3	-16.1
73	C	How do A- and H-bombs affect the health of the world?	35.1	39.8	13.4
74	P	How do boys think and feel differently than girls?	63.1	44.4	-29.6
75	C	Why are "mad" dogs killed?	19.3	19.4	0.5
76	P	How can one learn to get along with other people?	29.8	23.1	-22.5
77	P	Why should one eat a variety of foods?	7.0	12.0	71.4
78	C	Why must farmers have their cows tested before they can sell milk to the public?	5.3	24.1	354.7

TABLE I (continued)

SUMMARY OF THE RESULTS OF THE TWO ADMINISTRATIONS

Item	Personal or Community	Question	% , 1st Time	% , 2nd Time	% Gain or Loss
79	P	Can a mother "mark" her unborn child by her thoughts or emotions?	63.1	46.3	-26.6
80	P	Is a balanced diet expensive?	4.4	13.9	215.9
81	P	Are raw foods dangerous?	23.7	20.4	-13.9
82	C	How can one know whether what one reads and hears about health is true?	10.5	25.0	138.1
83	C	Is there a "world" health department?	11.4	25.0	119.2
84	P	Is fresh air really necessary?	15.8	15.7	-0.6
85	P	Should everyone be immunized against diphtheria?	13.1	19.4	48.1
86	C	How are germs spread from person to person?	9.6	19.4	102.1
87	P	What is immunization?	13.1	17.6	35.1
88	P	How and when should artificial respiration be given?	7.0	21.3	204.3
89	P	Is lightning dangerous to health?			
90	C	Why is it important to have a birth certificate?	28.0	18.5	-33.9
91	P	How can one splint a broken limb?	14.9	19.4	30.2
92	P	What are antibiotics and how are they used?	13.1	15.7	19.8
93	P	How and when should a person's pulse and temperature be taken?	15.8	18.5	17.1
94	C	Should peddlers of narcotic drugs be imprisoned?	33.3	25.9	-22.2
95	P	How can I make myself more interesting to others?	38.6	23.1	-40.2
96	C	Why do cities have classes to show new mothers how to care for their babies?	16.6	25.0	50.6
97	P	Is feeble-mindedness inherited?	45.6	25.9	-43.2
98	P	What are some symptoms of mental illness?	31.6	19.4	-38.6
99	P	Can mental illness be cured?	24.5	25.9	5.7
100	C	Should all sick immigrants be kept out of this country?	14.0	22.2	58.6
101	P	Can alcoholism be cured?	30.7	23.1	-24.8
102	P	Is it safe to leave food in opened tin cans for some time?	14.0	20.4	45.7
103	C	Can garbage safely be fed to hogs?	14.0	20.4	45.7

TABLE I (continued)

SUMMARY OF THE RESULTS OF THE TWO ADMINISTRATIONS

Item	Personal or Community	Question	%, 1st Time	%, 2nd Time	% Gain or Loss
104	P	Are x-rays harmful?	14.0	26.8	91.4
105	P	Does weather have an effect on one's health?	20.2	28.8	42.6
106	P	Will physical activity improve one's appearance?	28.9	19.4	-32.9
107	C	What effect might war have on the health of a country?	13.6	29.6	117.9
108	P	How much sleep should a high school pupil get every night?	49.1	27.7	-43.6
109	P	What first aid should be given to a person who has been poisoned?	25.4	23.1	-9.1
110	C	Should sick people be required by law to go to hospitals?	15.8	21.3	34.8
111	P	What causes high blood pressure?	31.6	38.8	22.8
112	P	How can water found on a hike or at a picnic spot be made safe to drink?	24.5	25.9	5.7
113	C	Would it be better to leave all health matters to private individuals and organizations rather than have tax-supported health departments?	6.1	21.3	249.2
114	P	Will alcohol cure snake-bite?	40.4	34.3	-15.1
115	P	What is the first-aid treatment for sunstroke? For heat prostration?	26.3	29.6	12.5
116	C	Is there any relationship between housing conditions and disease?	12.2	14.8	21.3
117	P	What three foods would make a completely balanced diet?	21.0	22.2	5.7
118	P	What is a virus?	21.0	19.4	-7.6
119	C	Why do some communities add fluorine compounds to their water supplies?	7.9	15.7	98.7
120	C	What diseases are inherited?	41.2	25.0	-39.3
121	P	Does candy give children "worms"?	40.4	35.2	-12.9
122	P	Should one eat the same foods in summer as in winter?	23.7	25.0	5.5
123	P	What causes hardening of the arteries?	25.4	32.4	27.5
124	P	Why do we get a "lump in the throat" from deep emotion?	61.4	49.1	-20.0
125	P	What is the difference between antiseptics and disinfectants?	28.0	42.6	52.1

--39.8%.	
Personal item of least interest, first time,	80
--4.4%.	
Personal item of least interest, second time,	77
--12.0%.	
Community item of least interest, first time,	2
--3.5%.	
Community item of least interest, second time,	67
--12.9%.	

The average gain in interest in the community health items was much greater than the average gain for the personal health items. This, of course, was a desired outcome. The average loss, in both categories, was very nearly the same. The percentage of community health items showing gain in interest was considerably larger than the percentage of personal items showing gain, and the percentage of personal items showing loss in interest was somewhat larger than the percentage of community items showing loss. On the whole, a greater interest in community health problems was shown at the second marking of the inventory.

When lists of the ten most interesting items were compiled for each of the two administrations of the inventory, eight items appeared on both lists, showing that for these items, pupil interest had not been satiated. Of the ten items of least interest in the first administration of the inventory, five were on community health, but none of these five appeared among the least interesting questions in the second administration. Three other items on community hygiene appear among the least interesting questions in the second administration, but the percentages of interest were much higher the second time and each of these three items actually registered an appreciable gain in interest. One outstanding addition to the list of most interesting items on the second administration was Item 73, the effect of A- and H-bombs on the health of the world. Increased concern for this problem is of great social significance. Other items on community health in which high interest was shown on both administrations of the inventory deal with smoke and smog control, the most destructive diseases in the world, and required premarital blood tests.

In summary, it can be said that a definite gain of interest in community health problems was shown by the group as a whole on the second administration of the inventory, both in number of items holding interest and in percentages of pupils interested in each problem. Although there are differences among the various questions, the total result supports the hypothesis that pupil interest in community health problems can be stimulated. Since interest in personal health questions remained high, it is important to note that the gain of interest in community health problems was not made at the expense of interest in personal health problems.

Implications

The gain of interest in community health problems confirms the desirability of including this topic in the high school instruction program. The study indicates that a teacher need not always confine himself to satisfying existing interest of his pupils, but can stimulate interest in areas which the experience of more mature minds has shown to be valuable. It is possible that lack of interest in a subject may indicate mere unfamiliarity with it, and that more information may create greater interest. There usually comes a point, of course, where mastery of basic facts about fundamental or mechanical processes will reduce interest; but the problems of community and world health are so far-reaching and so ever-changing that this point of diminishing returns is not likely to be reached at the high school level. For example, pupils can master the information about what constitutes an adequate diet, but the problem of making such a diet available to all the peoples of the world is not likely to be solved for many years to come. Not until our youth become aware that it is a problem will they bring their intellects and energies to bear upon its solution.

In biology or health courses, attention to community responsibilities for maintaining the health of individuals and the responsibilities of individuals for maintaining the health of the community as a whole can be a rewarding experience. Social studies courses have opportunities for including work on community health with emphasis on the reasons for community responsibility in this field and the economic aspects of public health. A study of the culture, political philosophy, and economic stability of countries with a low level of health might suggest the desirability of raising the level of health in the world at large as one means of influencing international relations.

This study also suggests several other topics for investigation. In a very broad and general way, it indicates that teachers in any area might find it helpful to survey the interests of their pupils from time to time, at least informally, as a guide to course content, organization, and emphasis. While pupil interests cannot be taken as a sole basis for instructional planning, such surveys might reveal areas which are worthy of consideration.

Since this study made no attempt to determine reasons for interest or lack of interest, an attempt to determine the relationship between knowledge and interests might be worthwhile. Does interest increase with acquisition of information? Or is a little knowledge of a given topic usually accompanied by little or no interest in that topic?

The final step, of course, would be a study of the relationship between interest and behavior. Will great interest in a topic influence behavior? For not until behavior has been changed can learning be said to have taken place.

REFERENCES

1. Byrd, O. E. "Health Problems for Course and Curriculum." The Research Quarterly of the American Association for Health, Physical Education, and Recreation. March, 1950.
2. Denver Public Schools. "The Twelve Items of Greatest Interest to Pupils in Grades Four to Twelve." The Health Interests of Children, 1947.
3. Lantagne, Joseph E. "Health Interests of 10,000 Secondary School Pupils?" The Research Quarterly, October, 1952.
4. Moon, Truman J., Mann, Paul B., and Otto, James H. Modern Biology. Henry Holt and Company, New York, New York; 1948.
5. Mooney, Ross L. Mooney Problem Check-List. 1950 Revision, Form H. The Psychological Corporation, New York, New York.
6. Southworth, Warren H., Latimer, Jean V., and Turner, C. E. "A Study of the Health Practices, Knowledge, Attitudes, and Interests of Senior High School Pupils." The Research Quarterly, May, 1944.

The National Science Foundation announces that it will continue to support the following programs during the summer of 1960 and the academic year 1960-61:

Undergraduate Research Participation,
Research Participation for Teacher Training,
Science Training for Secondary School Students.

Colleges, universities, and other educational institutions desiring to consider participation in these programs may receive full information by addressing:

Special Projects in Science Education,
Scientific Personnel and Education,
Washington 25, D. C.

RURAL-URBAN DIFFERENCES IN INTELLIGENCE *

IRVIN J. LEHMANN
Michigan State University

IT HAS LONG been recognized that there are cultural differences in intelligence test performance. Binet (2) removed certain items from his 1905 scale because he felt that their responses were influenced by home and/or school training. Eells (5), Havighurst (6), Terman (11), Sherman and Key (9), Asher (1), and Wheeler (12) have shown that there are differences in the intelligence test performance of children coming from different socio-economic strata.

Some investigators (4, 8, 10) feel that rural children score lower on intelligence tests because the items are based upon experiences common to the American urban culture. Davis (4) feels that there has not been sufficient research conducted to determine the proportion of items that favor urban over rural children. If rural-urban differences in intelligence test performance exist, the extent to which they are significant may depend, in part at least, upon the dissimilarity of the experiential background between rural and urban children. Boger (3) has shown that when rural children are given training in answering current intelligence test items, they have a much higher score than those children who were not trained. In other words, he feels that rural children have lower intelligence test scores because of the nature of the tests of mental ability being used.

Earlier investigations of rural-urban differences in intelligence have, for the most part, shown that urban children have a higher mean IQ and excel on the verbal portions of intelligence tests. These studies, however, were conducted as long as thirty years ago. In that time, with changes in our pattern of living, advances in the media of mass communication, and better rural schools, the experiences to which rural and urban children are subjected may have become so similar today that there is no longer any appreciable difference in intelligence as measured by our present tests of mental ability.

Statement of Problem

The major purpose of this study was to ascertain whether there is a significant difference between rural and urban children in IQ, MA, or responses to the individual items of Form L of the Revised Stanford-Binet.

Procedure

The subjects were forty-four rural and forty-four urban children from a city and an all-rural county in southeastern Wisconsin grouped according to age and sex. There were eight groups of eleven subjects each, differentiated as to whether they were first or sixth graders, boys or girls, and rural or urban. Certain criteria were established for choosing the subjects and for designating a school as rural.

1. Criteria for rural schools
 - a) Single-teacher.
 - b) Relatively distant from any large city.
2. Criteria for choosing the subjects
 - a) White.
 - b) American-born
 - c) Free from severe sensory defects, including vision, hearing, and speech.
 - d) Grade I children born between June 1 and December 31, 1948, and Grade VI children between June 1 and December 31, 1943.
 - e) The subjects had to live the latter half of their lives in the area in which they were residing at the time the study was made to be considered as rural or urban.

Of the 231 children found in the first and sixth grades in the rural schools of Midcounty¹, 79 met the criteria for eligibility and became the population from which the rural sample of 44 was randomly chosen.

Of the 1365 children found in the first and sixth grades in the urban schools of Midtown¹, 552 met the

*Footnotes will be found at the end of the article.

criteria for possible selection and became the population from which the urban sample of 44 was chosen in a random fashion.

Therefore, from the 1596 children enrolled in the first and sixth grades in the schools of an all-rural county and in a city in south-eastern Wisconsin, 631 met the criteria for possible selection and became the population from which the two samples of 44 were chosen in a random fashion.

The writer administered Form L of the Revised Stanford-Binet to the 88 subjects under relatively similar testing conditions. All the tests were scored by the writer according to the directions contained in the manual as well as in the supplementary guide (7). To insure accuracy in scoring, the test blanks were rescored one month after administration, and a further time just before the analysis of the data was undertaken.

Analysis of Results

Differences among the groups in IQ were tested for significance by the analysis of variance technique. Differences in MA among the groups were tested for significance by the analysis of covariance technique. The assumptions of homogeneity of variance, linearity of regression, and homogeneity of regression were tested by appropriate statistical techniques and found to be satisfied.

Differences between rural and urban children in their responses to the individual test items of Form L of the Revised Stanford-Binet were tested for significance by the use of chi-square. For the analysis of the responses to the individual test items, the first and sixth grade groups were tested separately.

A five per cent coefficient of risk was selected in advance of the analyses. Mean square ratios were considered significant if they equalled or exceeded the five per cent value in the F table. Throughout the study, differences are reported as either significant or not significant at the five per cent level of confidence.

Reference to Table I reveals that none of the interaction terms was significant. Since a linear hypothesis model was used in the study, the interaction terms were tested first, using the within term as the denominator in the F test. They were found to be not significant and hence were pooled with the within term to yield a "pooled error mean square," this pooled term being used as the denominator for the F test, the main effects. The denominator of the F test for the main effects is then 216.03 instead of 224.63. The only main effect found to be significant was the locale term. The mean IQ of the urban children is 111.7; the mean IQ of the rural children is 105.4. The results support the hypothesis of urban children having a significantly higher mean IQ than their rural counterparts.

Table II reveals that none of the main effects or interaction terms was significant at the five per cent level of confidence although the locale effect F of 3.77 approached the five per cent level, an F

value of 3.96 being required for significance at the five per cent level.

When the IQ scores were used, the difference in favor of the urban children is significant just beyond the five per cent level. When the MA scores were used, the difference in favor of the urban children approached the five per cent level. It is suggested that the reason for this discrepancy is that these rural children have higher mean CA's and lower mean MA's, whereas these urban children have lower mean CA's and higher mean MA's. If one were to represent the CA and MA scores pictorially, it would be seen that for the rural children, the CA scores are piled up at the high end of the scale and that the MA scores are piled up at the lower end. Conversely, the CA's of the urban children are gathered around the lower end while their MA's are clustered at the high end of the scale. When the MA scores of the rural and urban children were tested for significant differences by the analysis of covariance, the adjusted MA scores closely approached significance at the five per cent level. When the IQ scores were tested by analysis of variance, the fact that the IQ is the ratio of CA to MA resulted in the urban children having higher IQ's with the end result that the urban children have a significantly higher mean IQ than the rural children.

Numerous investigators have shown that urban children have a higher mean IQ than their rural counterparts. When the writer surveyed the literature on rural-urban differences in intelligence, he found that many of them did not determine whether the difference in favor of the urban children was significant. The writer chose nine such studies for a composite analysis and found the difference in favor of the urban children to be significant beyond the five per cent level. Table III reveals that when the various studies were treated as clusters from a common population and tested for difference by using cluster sampling techniques, the difference in favor of the urban children is significant beyond the 0.1 per cent level.

The second portion of this study was concerned with learning whether there is a significant difference between rural and urban children in their responses to the individual items of Form L of the Revised Stanford-Binet. Since the writer did not administer the complete test to every subject², it was necessary to analyze the responses for the first and sixth grade subjects separately. This was done by using the chi-square technique and applying Yates' correction. The results of the present study do not support the general belief that rural children do better on non-verbal items and that urban children excel on verbal items. For the first grade subjects, more rural than urban children responded correctly to 7 items whereas more urban than rural children responded correctly to 35 items. This difference is significant beyond the five per cent level. For the sixth grade subjects, more rural than urban children responded correctly to fourteen items while more urban than rural children responded to 35

TABLE I
ANALYSIS OF VARIANCE OF THE IQ SCORES

Sources of Variation	df	Mean Square
Age	1	90.01
Sex	1	10.92
Locale	1	871.92*
Age x Sex	1	27.29
Age x Locale	1	.29
Locale x Sex	1	8.28
Age x Locale x Sex	1	140.01
Error	80	224.63**

* Significant beyond the five per cent level.

** The error term was only used to test the interaction effects. When they were found to be not significant a "pooled error mean square" of 216.03 was used to test the main effects.

TABLE II
ANALYSIS OF VARIANCE OF THE ADJUSTED
MA SCORES

Sources of Variation	df	Mean Square
Age	1	135.43
Sex	1	.01
Locale	1	1062.57
Age x Locale	1	87.43
Age x Sex	1	29.41
Sex x Locale	1	62.11
Age x Locale x Sex	1	250.56
Error	79	281.13

TABLE III
COMPARISON OF THE MEAN IQ FOR RURAL AND URBAN
CHILDREN WHEN TREATED AS CLUSTERS*

Study	N_R	N_U	\bar{X}_R	\bar{X}_U	\bar{X}
1	818	1327	97.6	103.2	100.4
2	29	81	105.0	109.0	107.0
3	30	84	103.0	107.0	105.0
4	115	304	108.2	108.2	108.2
5	2211	591	102.9	101.3	102.1
6	864	422	95.4	105.8	100.6
7	180	188	100.6	106.3	103.5
8	546	410	95.4	105.8	100.6
9	153	154	95.7	107.9	101.8
Mean			100.4	106.1	103.2

* The formula used to calculate the denominator for use in the t-test was

$$s_{\bar{X}}^2 = \frac{m}{m-1} \frac{\sum_{i=1}^m N_c^2 (\bar{x}_i - \bar{X})^2}{\left(\sum_{i=1}^m N_c \right)^2}$$

where m = number of clusters or studies
 n_c = number of individuals in the i^{th} cluster.
 \bar{x}_i = mean IQ for the N_i individuals in the i^{th} cluster.
 \bar{X} = mean IQ of the sample.

$$t = \frac{\bar{X}_U - \bar{X}_R}{s_{\bar{X}}}, \quad t = \frac{106.1 - 100.4}{\sqrt{\frac{9}{8} \frac{(57,686,129.30)}{(72,199,009.00)}}} = 6.18$$

For 8 df, $\Pr(t \geq 5.04) = .001$

TABLE IV
COMPARISON OF THE MEANS AND STANDARD DEVIATIONS OF IQ'S OF RURAL AND
URBAN CHILDREN IN THIS STUDY WITH THE RURAL AND URBAN CHILDREN
IN THE 1937 STANDARDIZATION GROUP

	Present Study			1937 Standardization Group			
	N	Mean	S. D.	N	Mean	S. D.	t
Urban Children	44	111.7	13.8	864	105.8	14.7	2.61*
Rural Children	44	105.4	15.3	422	95.4	15.5	2.95*

*Significant beyond the one per cent level.

items correctly. This difference is significant beyond the five per cent level. Although the urban children in both the first and sixth grades responded correctly to more items than their rural counterparts (significant beyond the five per cent level), insofar as the analysis of the individual test items is concerned, there is no significant difference in the number of correct responses to each of the test items between rural and urban children. It is suggested that the reason for findings such as these is due to the low power of chi-square for so few cases ($n = 11$ for each group). The fact that more urban than rural children responded correctly to more test items is reflected in the urban children having a higher mean MA score.

Previous research has shown that urban children have a higher mean IQ than their rural counterparts. When some of these studies were treated as clusters, it was found that the difference in IQ in favor of the urban children is significant beyond the 0.1 per cent level. The present study has shown similar findings. A point of interest which might be raised is "Are children today less or more intelligent than those two decades ago?" Although the subjects of the present study were restricted to a very limited population in southeastern Wisconsin, Table IV shows that the subjects of the present study have higher mean IQ's and smaller standard deviations than the rural and urban children used in the 1937 standardization of Form L of the Revised Stanford-Binet. Lest one state with any degree of emphasis that the rural and urban children in the present study are more intelligent than those rural and urban children used in 1937, it should be remembered that the subjects of the present study were restricted geographically, and were not stratified according to socio-economic status; also the criteria for defining rural and urban were different in the two studies. It should also be recalled that Wisconsin children were not included in the 1937 standardization group. The results of the present study are only suggestive and further research is needed before any definite conclusions may be reached.

Conclusions

The findings of the present study support those of previous investigations on rural-urban differences in IQ. Even though different populations were used, and different tests of mental ability were employed between this and some of the other studies, the results demonstrated that urban children have a higher mean IQ.

The analysis of the number of correct responses to the individual test items do not support conclusively the findings of previous studies which have shown that rural children do better on non-verbal items. The present study has shown that there is no significant difference between rural and urban children in the number of correct responses to the individual items of Form L of the Revised Stanford

-Binet. However, because of the low power of chi-square in the present study, the findings are only suggestive. Further research should be conducted to determine whether previous findings on the superior performance of urban children on verbal-type items still prevails.

The findings on the comparison of the mean IQ's of the subjects of the present study with those of the 1937 standardization group suggest that the differences between the two studies may be accounted for, in part at least, by changes in our social and economic ways of living which might have resulted in better physical and mental health. It is also possible that present-day children, because of their familiarity with tests, are able to obtain higher scores because they know what is expected of them.

The findings of the present study suggest the need of a wide testing program in order to learn whether the Revised Stanford-Binet is in need of further revision.

Summary

The present study has sought to test the hypotheses that 1) there is no significant difference in IQ and MA between rural and urban children, and 2) there is no significant difference between rural and urban children in their responses to the individual test items of Form L of the Revised Stanford-Binet.

To test these hypotheses, 44 rural and 44 urban children were randomly selected from a city and from an all-rural county in southeastern Wisconsin and were given Form L of the Revised Stanford-Binet by the writer.

Differences in IQ among the groups were tested for significance by the analysis of variance technique and it was found that the urban children have a significantly higher mean IQ. Differences among the groups in MA were tested for significance by analysis of covariance (CA was the covariate) and it was found that although urban children have a higher mean MA, the difference just approaches the five per cent level of significance.

Differences between rural and urban children in their responses to the individual test items were tested by chi-square and it was found that there is no significant difference. When the overall differences in number of correct responses to the individual test items were tested, it was found that for both the first and sixth grade subjects, more urban than rural children responded correctly to more test items, the difference being significant beyond the five per cent level.

FOOTNOTES

- * This article is based upon part of an unpublished Ph.D. dissertation completed at the University of Wisconsin, 1957, under the direction of Prof. Chester W. Harris.
- 1. The rural county is referred to as Midcounty, the urban area as Midtown.

2. The first grade subjects began the test at Year VI, the sixth grade subjects began the test at Year XI.

REFERENCES

1. Asher, E. J. "The Inadequacy of Current Intelligence Tests for Testing Kentucky Mountain Children," Journal of Genetic Psychology, 46, (June, 1935), pp. 480-486.
2. Binet, Alfred and Simon, Th. The Development of Intelligence in Children, translated by Elizabeth S. Kite, Baltimore: Williams and Wilkins Co., 1916.
3. Boger, Jack H. "An Experimental Study of the Effects of Perceptual Training on Group IQ Scores of Elementary Pupils in Rural Ungraded Schools," Journal of Educational Research, 46, (September, 1952), pp. 43-52.
4. Davis, Allison. Social Class Influences Upon Learning, Cambridge: Harvard University Press, 1948.
5. Eells, Kenneth W. and others. Intelligence and Cultural Differences, Chicago: University of Chicago Press, 1951.
6. Havighurst, Robert J. and Janke, Leota L. "Relations Between Ability and Social Status in a Midwestern Community," Journal of Educational Psychology, 35, (September, 1944), pp. 357-368.
7. Pinter, Rudolph and others. "Supplementary Guide for the Revised Stanford-Binet Scale (Form L)," Applied Psychology Monographs, No. 3, 1944.
8. Shepard, Eugene L. "Measurement of Certain Non-Verbal Abilities of Urban and Rural Children," Journal of Educational Psychology, 33, (September, 1942), pp. 458-462.
9. Sherman, Mandell and Key, Cora B. "The Intelligence of Isolated Mountain Children," Child Development, 4, (December, 1923), pp. 279-290.
10. Shimberg, M. E. "An Investigation into the Validity of Norms with Special Reference to Urban and Rural Groups," Archives of Psychology, No. 104, 1929.
11. Terman, Lewis M. Measurement of Intelligence, Boston: Houghton-Mifflin Co., 1916.
12. Wheeler, Lester R. "A Comparative Study of East Tennessee Mountain Children," Journal of Educational Psychology, 33, (May, 1942), pp. 321-334.

COLLEGE STUDENTS' ABILITY TO EVALUATE THEIR PERFORMANCE ON OBJECTIVE TESTS

EARL W. KOOKER, North Texas State College
CHESTER S. WILLIAMS, University of Oklahoma

RECENT DISCUSSIONS in the literature concerned with student evaluation emphasize the desirability of allowing the student to participate in the evaluation process. Most of these discussions have either been philosophical in nature or have dealt with the apparent consistency of this practice with certain psychological principles (3). To allow the student to participate in evaluating his own performance is held to be more democratic. It is further reasoned that the ability to look realistically at the results of one's own endeavors is an important part of the problem-solving process, and if the student is to develop problem-solving ability, he must have some direct experience in evaluating his own progress.

In spite of this accent in theoretical discussions, there seems to be a scarcity of research bearing directly on the use of this technique in school situations. As Russell points out in his review, there is more lip service than practice and there are many unanswered questions concerning the usefulness and practicality of self-evaluation (7). A few reports have been published describing the use and advantages of self-evaluation (1, 8). Some researchers have been concerned with the accuracy of student estimates of their own performance. Bear and Levin investigated awareness of vocabulary size among college students and its relation to class standing and sex. They concluded that a student's estimate of performance is an uncertain guide to his actual performance, and that the accuracy of his estimate is related to class standing but not to the sex of the individual (2, 6). That several variables are related to this accuracy has been demonstrated by Brim, Gilinsky and Sumner (4, 5, 9). Among these variables are past performance, IQ, and the group with which the individual is comparing himself. Young found that college students predicted their academic success more accurately than did college counselors (10).

It is frequently implied that students should participate in evaluating themselves even in areas in which the criteria are rather nebulously stated, such as progress toward broad behavioral goals. If students can do this, it certainly must be assumed that they would be able to estimate the accuracy of

their responses on more objectively stated tasks. In the light of the above research, it would seem that this assumption needs further investigation.

This study seeks to discover whether college students can accurately assess their own performance on objective test items and whether this ability changes over a semester's duration. Specifically it attempts to answer the following questions: 1) What percentage of the test items for which students say they are certain of the answers do they actually answer correctly? 2) Do they answer a greater percentage of these items correctly than they do items for which they say they are doubtful of the answers or are guessing on the answers? 3) Does the percentage of the accurately answered certain questions change significantly over a semester's duration?

Subjects

The subjects used in this investigation were college students in nine classes in the Department of Education and Psychology at North Texas State College. The classes employed ranged from first semester freshmen through the doctoral level. Included were classes in 1) general psychology (two sections, mostly freshmen and sophomores with a total N of 50); 2) beginning experimental psychology (mostly freshmen and sophomores with an N of 17); 3) child psychology (two sections, mostly sophomores and juniors with a total N of 50); 4) child psychology (mostly juniors and seniors with an N of 15); 5) individual intelligence testing (mostly juniors and seniors with an N of 21); 6) social foundations of education (two sections, all doctoral students with a total N of 32). In the remainder of this paper, the groups will be referred to by these number designations.

Procedure

In each class, students were asked to evaluate how confident they felt about their answer to each test question. If they were certain of their answer, they were asked to place a "C" beside the question number; if they were doubtful, they were to use a

"D" and if they were guessing, a "G". The tests, except in one group, included both multiple-choice and true-false questions. In this one instance, only true-false questions were used. Most of the multiple-choice items were four-choice items.

The examinations employed were the regular teacher-made examinations used for evaluation purposes. The students were told that the C, D, and G had nothing to do with the grades they would receive. It was explained that the instructors were doing some research in which they were attempting to discover how well students could estimate their performance on such tests.

In all but one group, three tests were given with each test covering about one third of the semester's work. The number of questions varied somewhat from test to test with a slightly larger number of true-false than multiple-choice questions. In each case, the first two tests were graded and handed back to the students, and questions which they had regarding the corrections were discussed. In group 6), only one true-false test was given.

The accuracy of the student's judgments on the first test was determined by calculating the percentage of "C" questions answered correctly and the percentage of "D" plus "G" questions. The "D" and "G" questions were combined because of the relatively small number of "G" questions. The multiple-choice and true-false items were analyzed separately. For the "C" questions, similar percentages were calculated for the final test in each group and the significance of the difference between initial and final means were tested by the use of t .

Results and Discussions

The M percentage of items answered correctly in the certain and doubtful-guessing categories for each type of item and the associated SD's are summarized in the left half of Table I. These data are from the initial tests in each of the six groups. In the right half of the table is summarized the tests of significance of the difference between the M percentage of the certain items answered correctly for the first and last of the three tests given.

An examination of Table I reveals that within all groups, for both types of items, a greater percentage of the "C" items were answered correctly than were the "D-G" items. In all five groups, in which both types of tests were given, a greater percentage of the true-false items marked "C" were answered correctly than were the multiple-choice items marked "C". Further analysis of difference by type of item reveals that within all groups, a greater percentage of the true-false "D-G" items were answered correctly than were multiple-choice items for which the subjects had marked "D-G". Summarizing the data across groups, the Mdn of the percentage answered correctly by type of item and category were as follows: 1) 84 for true-false "C" items; 2) 60 for true-false "D-G" items; 3) 76 for multiple-choice "C" items; 4) 55 for mul-

multiple-choice "D-G" items. These results suggest that the college students used in this study were able to select the items for which they knew the answers with some degree of accuracy. They displayed some ability to discriminate between those items they knew and those they did not know as evidenced by the consistently lower percentage of "D-G" items answered correctly as compared to the "C" items.

As would be expected, the subjects missed more of the multiple-choice items placed in the "D-G" category than they did true-false items. This should occur if they were really doubtful or guessing, since they had only two alternatives on the true-false items and four on the multiple-choice items.

As for the variabilities within groups, in five of the six groups the SD's were smaller for the "C" items on the true-false tests than for the "D-G" items. In four of the five groups, the same trend was evident for the multiple-choice items. The Mdn of the SD's for the true-false "C" items was 10.5, for the true-false "D-G" items 16.4, for the multiple-choice "C" items 16.0, and for the multiple-choice "D-G" items, 20.1. These results suggest that the "D-G" category was more vaguely conceptualized by the subjects and that there is more variability in estimating performance on multiple-choice items than on true-false items.

The changes in the M's for the "C" items on the true-false tests over the semester were not consistent. In one instance there was a significant increase in the M percentage answered correctly, in another a significant decrease and in the remaining three groups there was no significant change. On the multiple-choice tests the M's on the final test were all higher than on the initial tests. In three groups the change was statistically significant but in the remaining two the change was significant at only the .11 and .15 level.

It will be recalled that the subjects had the opportunity to discuss any questions on the first two tests which were corrected and handed back so that the subjects might discover why they missed items for which they felt they knew the answer. In spite of this, on the true-false tests the subjects did not consistently demonstrate any increase in the ability to answer correctly items for which they said they were certain of the answer. There is a suggestion of such an increase on the multiple-choice items. Since all of the tests used had been given to several other classes and discussed with them and an effort made to eliminate ambiguous items, this factor should not account for the differences between multiple-choice and true-false items found here. It may, of course, be that there is more to know about how to take a multiple-choice test or this particular variation of a multiple-choice test may have been relatively newer to more students than was the true-false form of testing.

The procedures used in this study were such as to place the student in as nonthreatening a situation as possible, for he was assured that his self-eval-

TABLE I

MEASURES OF CENTRAL TENDENCY AND VARIABILITY FOR THE PERCENTAGE OF EACH TYPE OF ITEM ANSWERED CORRECTLY IN EACH CATEGORY BY THE SIX GROUPS AND THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE MEANS OF INITIAL AND FINAL TESTS FOR THE CERTAIN ITEMS

Group	Type of Item	How Marked	M	SD	$M_2 - M_1$	SE D	df	t	P
1)	T-F	C	78	12.4	1.1	1.9	47	.58	-
		D-G	57	15.2					
	M-C	C	73	20.6	9.7	2.8	49	3.49	.001
		D-G	37	17.2					
2)	T-F	C	86	10.4	-11.5	2.3	16	-5.0	.001
		D-G	59	17.2					
	M-C	C	76	16.0	6.1	3.5	16	1.7	.11
		D-G	47	20.1					
3)	T-F	C	82	13.4	6.4	1.8	49	3.56	.001
		D-G	61	16.3					
	M-C	C	79	15.5	7.2	1.6	49	4.50	.001
		D-G	55	20.2					
4)	T-F	C	87	9.7	1.6	3.0	14	.53	-
		D-G	64	16.5					
	M-C	C	81	16.0	5.6	3.5	14	1.60	.15
		D-G	61	17.2					
5)	T-F	C	87	10.5	1.2	1.9	20	.62	-
		D-G	62	24.3					
	M-C	C	75	13.0	6.9	2.9	20	2.38	.03
		D-G	60	24.2					
6)	T-F	C	74	10.2					
		D-G	58	7.6					

uation would in no way affect his grade. Whether a student would be as objective in evaluating his own performance in situations in which his grade would be dependent upon his self-evaluation needs further investigation. Also, estimating one's performance on individual items is probably not perceived by the student in the same way as giving himself a grade.

Summary and Conclusion

Six groups of college students were asked on a true-false and multiple-choice test to indicate whether they were certain, doubtful or guessing in their responses. The percentage of correct responses was determined for the "certain" items and for "doubtful" plus "guessing" items. The *Ms* and *SD*'s for these percentages were determined for each group. In all groups a higher percentage of the "C" items were answered correctly than were the "D-G" items and a higher percentage of the true-false items were answered correctly than were the multiple-choice items. The sigmas tended to be higher for the "D-G" items than for the "C" items and higher for the multiple choice items than for the true-false items.

When the *M* percentage correct of the "C" on the first test was compared with this *M* on the last test of the semester, there was no consistent difference for the true-false items. There was some evidence of a significant change on the multiple-choice items. It was concluded that college students used as subjects do possess some ability to select items to which they know the answer and to differentiate them from items for which they are doubtful or guessing on the answer. This ability as measured by the percentage of "certain" items they get right does not change significantly over a semester's

duration for true-false items, but does tend to increase for multiple-choice items.

REFERENCES

1. Alluras, L. "Experiment in Self-Evaluation," *Nations Schools*, 1942, 30, 25.
2. Bear, R. N. and Odbert, H. S. "Insight of Older Pupils into Their Knowledge of Word Meanings," *School Review*, 1941, 49, 754-765.
3. Blair, G. M., Jones, R. S. and Simpson, R. H. *Educational Psychology*. (New York: The Macmillan Co., 1954).
4. Brim, O. G., Jr. "College Grades and Self-Estimates of Intelligence," *Journal of Educational Psychology*, 1954, 45, 477-484.
5. Gilinsky, A. S. "Relative Self-Estimate and Level of Aspiration," *Journal of Experimental Psychology*, 1949, 39, 256-259.
6. Levin, Janice and Stacey, C. L. "Awareness of Vocabulary Size: Its Relation to Class Standing and Sex Differences," *Journal of Educational Psychology*, 1951, 42, 174-184.
7. Russell, D. H. "What Does Research Say About Self-Evaluation?" *Journal of Educational Research*, 1953, 46, 561-573.
8. Rockowitz, M. "Know Thyself: Techniques in Pupil Self-Evaluation," *High Points*, 1951, 33, 50-56.
9. Sumner, F. C. and Johnson, E. E. "Sex Differences in Levels of Aspiration and in Self-Estimates of Performance in a Classroom Situation," *Journal of Psychology*, 1949, 27, 483-490.
10. Young, F. C. "College Freshmen Judge Their Own Scholastic Success," *Personnel and Guidance Journal*, 1954, 32, 399-403.

MEASUREMENT OF TEACHER MERIT FOR SALARY PURPOSES *

WILLIAM A. McCALL**, GERTRUDE R. KRAUSE
Coral Gables, Florida

Total Growth and Total Teacher Merit

Seventy-three sixth-grade rural and city, Negro and white teachers in the schools of North Carolina participated in the investigation here reported. Since the criterion of merit adopted for the study was the amount of good all-round growth each teacher produced in his or her class, the initial status of the pupils in September and the final status in May of what is sometimes called the nine R's--Reading, Riting, Rithmetic, Research, Reasoning, Reporting, Relationship of persons, Recreation, and Responsible work skills--was measured.

The growths each teacher produced in the above nine areas were combined, with acceptable weightings according to importance, into a single criterion-of-merit score.

Since these growths were found to be affected by IQ, pupil drive, educativeness of home and community, class size, and regularity of attendance, each of these was measured, the amount of influence determined, and proper allowance made for each. Thus was secured a criterion score of merit for each teacher. These final corrected scores showed a surprisingly wide range of teacher efficiency, the poorest rating being twenty and the highest eighty-eight.

Obviously it is impractical to employ such an extensive program of testing and estimation whenever it is desired to discover a teacher's merit, but the procedure is not without merit. Accordingly an attempt was made to find out whether there were any more easily measured characteristics of teachers which correlated with the criterion. In general, the findings agreed with those previously discovered in the similar but more limited investigation by Crabbs¹. McCall obtained, briefly stated, the following findings:

1. Superintendents, supervisors, principals and colleagues tended to rate good teachers low and poor teachers high. The correlation for ratings by principals and ratings by peers was the same, $-.11$. College and university professors were no better judges of merit, for the marks they assigned these

teachers when they were in training also showed a negative correlation with the criterion of $-.11$.

2. The only persons in the school system who were found to be professionally competent to judge the worth of teachers were their sixth-grade pupils ($.36$) and the teachers themselves ($.39$) when giving a confidential self-rating.

3. Training, an almost universally employed basis for evaluating teacher merit and fixing salaries, was somewhat better than drawing shuffled names out of a hat ($.13$).

4. Years of service, usually referred to as experience, another customary basis for determining the worth and salary of teachers, showed a zero ($-.04$) correlation with merit. This finding is supported by the further fact that young teachers averaged higher criterion scores, or over-all growth in pupils, than did older teachers.

5. The teacher's knowledge of the subject-matter showed no correlation ($-.06$) with efficiency. Perhaps those who know the most cannot resist the temptation to lecture, and those who learn with their pupils are better teachers.

6. Marital status, often used in times past to judge the desirability of employing teachers, showed some correlation, but it has been used in the wrong direction, for married teachers were superior to unmarried ones. This does not indicate, of course, that boards of education should set up marriage bureaus in order to improve teaching, for quite possibly men have preferred for wives those who are good teachers, the dynamic factors being physical and emotional normality or something else not yet identified. Emotional normality was measured and showed a positive correlation ($.30$) with merit.

The worth of certain less-commonly used methods of measuring merit was explored with the following results:

1. The Morris Trait Index, which purports to measure the degree of success with which a teacher can cope with both usual and unusual classroom situations, showed no relationship with the teachers' criterion scores.

* Footnotes will be found at the end of this article.

2. The McCall Professional Knowledge Test consists of many items, each one listing five pedagogical terms, one of which does not belong with the other four. Teachers taking the test must divine what the point of view is and identify the unacceptable term. This test was found to have a small positive relationship of .17 with the teachers' merit.

3. The McCall Political Tolerance Test, which shows where a teacher would score on a conservative-liberal continuum, showed a similar positive relationship (.18) between teacher efficiency and the degree of liberality of the teacher's political beliefs.

4. The McCall-Herring-Loftus School Practices Questionnaire, measuring the degree with which a teacher subscribes to progressive education, showed a positive relationship of .15 in favor of progressive education.

5. McCall attempted to determine whether the teacher's knowledge of various popular sports and the frequency of his participation in sports had any relation to the teacher's criterion scores. Knowledge of sports showed insignificant, although positive relationship, but participation showed a negative correlation of -.25.

6. The McCall Multimental Test, regarded as a measure of general mental ability or intelligence, showed a small positive correlation of .09 with the teachers' criterion scores and the McCall Word Knowledge Test proved equally low (.10).

7. Scores on tests and questionnaires in various areas which are considered to have a bearing on teacher efficiency proved to have even lower or no relationship to teacher efficiency. These tests were a questionnaire on social participation, measuring the amount and variety of a teacher's participation in civic, professional, and cultural activities; and a questionnaire on job participation which measured the variety of jobs worked at for pay since the teacher's graduation from elementary school.

8. Of all the measures used, the one which proved to have the highest correlation with teacher merit was the McCall-Herring Personality Measure, when used as a rating-by-pupils device. The scale consists of five items-- "is clean", "has good manners", "keeps temper", "is kind", and "is good citizen." Relationships for each item were computed separately and the correlations ranged from .22 to .39, all being positive. A comparison of these results with the results obtained from ratings of teachers by their peers, principals, and supervisors, indicates quite clearly that a teacher's pupils are far better judges of a teacher's merit than are professionally trained adults.

It is a well-established fact that any error of measurement present in the criterion scores and any other sets of measures with which they are correlated reduces the correlation towards zero in proportion to the size of the error. All competent persons will readily agree that there must be considerable error in the criterion scores. All measurements of complex human functions fall far short of perfection, and teacher merit is an ex-

ceedingly complex function.

Since the error in the criterion scores is unknown, we can only estimate a correction based on experience with mental measurements in general. Most of the other scores used in this study, especially ratings of the teachers by the children, were similarly affected by error. In the opinion of the writers, all coefficients of correlation reported in this study should be nearly doubled. Thus, when the study found a relationship between teacher personality and pupil growth of .39, the true relationship is much closer, perhaps about .70. Similarly all positive coefficients of correlations should be more positive than those reported and all negative coefficients more negative.

Personality Growth and Teacher Efficiency

Krause², whose primary interest is in guidance and counseling, made a further study of that portion of the pupils' measurements dealing with personality, together with the various measurements of their teachers, as gathered by McCall. The data were analyzed and evaluated in an attempt to establish other significant relationships. The following is a report on the results of this project.

The personality growth measure of the pupils had been obtained by having each child rate his classmates, at the beginning and end of the school year, using the McCall-Herring Personality Measure. This is the same instrument, described above, which the children employed in rating their teachers.

A statistical analysis of this material revealed the existence of an error of measurement. This was an end-error which penalized those classes with high initial scores. A correction was made for this error of measurement.

Since it was not known whether pupils' rate of growth in personality was affected by intelligence, this matter was further studied. It seemed from the data that classes with average IQ's below the ninety-six level were limited in ability to grow in personality. Accordingly, in fairness to the teacher, each class's average growth in personality was corrected for IQ.

When measures of the characteristics of the seventy-three teachers were correlated with this twice-refined measure of pupil personality growth, the following results were secured:

1. The most significant positive relationship found was between pupil personality growth and the pupils' ratings of their teachers' personalities.

2. Classes taught by teachers whose average college grades were below ninety per cent achieved good growth. Classes whose teachers averaged above ninety per cent made very small gains.

3. Classes taught by teachers whose ages ranged from forty-nine to fifty-six showed low growth compared with the classes taught by teachers whose ages ranged from twenty-nine to forty-two.

4. Classes taught by teachers whose experience

in teaching ranged from twenty to thirty-one years showed relatively little growth as compared with classes with less experienced teachers.

5. With the exception of teachers of very superior mentality, there appeared to be an inverse relationship between teachers' mental ages and pupil personality growth, the growth steadily decreasing as the teachers' mental ages increased, although these results were not sufficiently reliable to be conclusive.

6. Insignificant or no relationships appeared when pupil personality growth was correlated with teachers' marital status, number of offspring, years of training, political tolerance and educational progressiveness.

Other findings of interest were:

1. Pupils rated their teachers substantially lower in temper control than in any other of the five personality characteristics.

2. Teachers were rated by their pupils shortly after the beginning of the school year and at the end. Comparisons of these ratings showed that fifty per cent of the teachers fell in the estimation of their pupils during the school year, thirty-six per cent rose, and fourteen per cent remained the same.

Educational Implications of Both Studies

One of the important findings of McCall's study was the relatively close relationship between teacher personality and pupils' comprehensive growth. While it is true that causal connections cannot be inferred from correlations, the correlations found in these studies, when reasonably interpreted, and in the absence of conflicting evidence, appear to justify the following suggestions:

1. Teacher training institutions might attempt to discover what makes teachers good and attempt to incorporate these qualities in prospective teachers.

2. More attention should be given to selecting or developing teacher personalities which are helpful to pupils.

3. Experience as currently evaluated should be replaced as a basis for fixing salaries by a more defensible criterion. This must, of course, be accomplished by gradual change.

4. Ratings by superintendents, supervisors, principals, colleagues, or professors in teachers' colleges should not be accepted as the sole or valid criteria until persons in these positions have been professionally re-educated for this responsibility. Most administrators will feel that this finding supports their contention that they should not be asked to rate teachers for salary purposes.

5. Teachers' colleges should not expect teachers to become more effective in their profession merely as a result of acquiring more knowledge of the subject being taught. There is probably some critical minimum and this may be different for different types of subjects and schools.

6. The findings of McCall and Crabbs would appear to invalidate a large number of researches which are based on the assumption that the worth of a teacher can be validly judged by superior officials, or based upon training, experience, and knowledge of the subject taught. This means that future research, employing a merit criterion, should use either pupils' growth as a criterion or some combination of teacher characteristics really known to correlate with merit.

FOOTNOTES

- * The following persons and groups owe a debt of gratitude to the teachers, principals, supervisors and superintendents who were subjects in this investigation and who voluntarily agreed to risk possible personal embarrassment in order that this vital study might be made: all teachers, all pupils, the authors, the State Education Commission of North Carolina; its Executive Secretary, Dr. W. H. Plemmons, the State Board of Education of North Carolina and its Executive Secretary for the study, Dr. Allan S. Hurlburt. The state of North Carolina should be credited with support of this study. The State Education Commission, the State Board of Education and their representatives accepted and published the factual findings. The authors are responsible for the formulation of conclusions and interpretations in this article.
- ** At the request of the State of North Carolina, William A. McCall planned, directed and described a research project to determine the worth of existing methods and proposed methods of measuring teacher merit for salary purposes. A complete report appears in: William A. McCall, Measurement of Teacher Merit, State Superintendent of Public Instruction, Raleigh, N. C., Publication No. 284, 1952.
- 1. Crabbs, Lelah Mae, Measuring Efficiency of Teaching and Supervision, Teachers College, Columbia University, N. Y., 1925.
- 2. Krause, Gertrude R., The Relation of Certain Characteristics of Teachers to Personality Growth of Their Pupils, Master's Thesis, University of Miami, Coral Gables, Fla., 1956.

A STUDY OF THE EFFECTIVENESS OF A SUMMER REMEDIAL COURSE IN ENGLISH FOR COLLEGE FRESHMEN*

MARTIN L. ZEIGLER, LOUIS M. HERMAN
Pennsylvania State University

ONE OF THE COMMON problems facing many colleges and universities is the inadequacy of entering freshmen in English grammar and composition. At the Pennsylvania State University a remedial noncredit course--called Composition Zero--attempts to prepare students scoring below a critical score on an English proficiency examination (1) for subsequent credit courses in English. A previous study by Bowman, Zeigler, and Bernreuter (2) had shown that students who scored poorly on the proficiency test were likely to do poorly in the first credit course in English--Composition One.

The Composition Zero course is offered to prospective freshmen the summer before their formal enrollment, as well as during the regular fall semester. Students electing the Composition Zero course in the summer and passing it have an advantage in that they may pursue their regular English course work without delay in the fall. Thus, if the summer course is effective, it would be worthwhile to encourage prospective freshmen to elect it.

The present study attempts to determine if there is any significant difference in achievement in the subsequent Composition One course between students electing Composition Zero in the summer and a matched group of students electing it in the fall. There is a semester's gap in Composition One between these two groups, of course, since the latter group does not take it until the spring semester. An attempt is also made to measure the efficacy of both review courses by comparing grades of these review groups in Composition One with grades of a group of students matched for general abilities, but who had been exempted from the review course because of sufficiently high scores on the English proficiency examination.

Subjects and Procedure

All subjects were drawn from the entering freshman class of the fall of 1956. Three groups comprised the study. Group I (the summer-review

group) consisted of 25 out of a total of 30 students scoring below the cut-off score of 49 on the English proficiency examination and electing to take the required review course in the summer of 1956. The remaining five students were dropped from the study for various reasons, such as unavailability of grades and not continuing with Composition One in the fall. Group II (the fall-review group) consisted of 25 matched students falling below the cut-off score and electing to take the required review course in the fall of 1956. Group III (the no-review group) consisted of 25 matched students scoring above the cut-off score and not required to take the review course. These students took Composition One in the fall of 1956.

Members of Groups II and III were individually matched with members of Group I. Group II was matched with Group I on high school standing, general aptitude (3) and English proficiency; Group III was matched with Group I on high school standing and general aptitude only, since all members of Group III scored above the critical score on the English test and all members of Group I scored below it.

The criteria for matching on these various measures were:

- (a) high school rank: identical quintiles;
- (b) general ability: four-part-total scores on the aptitude test differing by no more than four points;
- (c) English proficiency: raw scores on the English proficiency examination differing by no more than five points.

Table I illustrates the overall results obtained in the matching procedures. Differences between any of the groups on the various measures were not statistically significant, with the exception of the no-review group's greater superiority to the two other groups on the English proficiency examination. This, of course, was a variable purposely introduced into the study.

* Footnotes will be found at the end of the article.

TABLE I
MATCHING OF THE THREE GROUPS ACCORDING TO THE
VARIOUS INDICATED MEASURES

Group	English Placement Test		Aptitude Test		High School Standing Quintiles	
	Raw Scores ^a		Four Part Total			
	Mean	S. D.	Mean	S. D.	Mean	S. D.
Summer Review	41.40	6.12	106.56	14.75	2.12	.65
Fall Review	41.48	5.59	107.12	13.81	2.12	.65
No Review	56.64	5.91	107.12	14.00	2.12	.65

^a Cut-off score was 48 or below.

TABLE II
GRADES OBTAINED BY EACH OF THE THREE GROUPS
OF 25 STUDENTS IN ENGLISH COMPOSITION ONE

Group	Number Receiving Each Letter Grade				
	F	D	C	B	A
Summer Review	10	7	7	1	0
Fall Review	4	13	8	0	0
No Review	3	9	12	1	0

Results and Conclusions

The three groups were compared on grades obtained in the English Composition One courses. Members of the groups were not necessarily in the same English One course section, but were distributed among various sections. The grades obtained by the three groups in Composition One are shown in Table II.

It is seen that grades for all three groups are relatively poor; in both the summer-review and fall-review groups, 68 per cent received a "D" or lower, and in the no-review group, 48 per cent were so graded. Although some differences in distribution of grades among the three groups is apparent, overall differences were not significant. The chi-square value for grade differences between the summer-review group and the fall-review group was 4.41 and for differences between the summer-review and no-review group 5.34. Neither of these values reached the previously established significance level of .05, since with two degrees of freedom a chi-square value of 5.99 is required.²

It appears then that grades in Composition One of students having taken the summer-review course do not differ significantly from grades of students having taken the fall-review course. Both review courses seem effective in raising the English proficiency of their students close to the level of ex-

empted students of about equal general ability, but of initially greater proficiency in English.

FOOTNOTES

1. Grateful acknowledgement is made to Margaret Ryan who helped in the matching procedures and statistical analysis.
2. Since there were no A's and only two B's for all three groups, only three grade categories were compared in the chi-square analysis -- "F", "D", and "C or above."

REFERENCES

1. Bowman, J. S., Patterson, H. D., and Sandin, E. V. The Pennsylvania College Placement Test in English, The Pennsylvania State University, 1949.
2. Bowman, J. S., Zeigler, M. L., and Bernreuter, R. G. The Relationship Between English Proficiency and Academic Performance in College. The Pennsylvania State University, 1954.
3. Moore, B. F., and Castore, G. F. The Pennsylvania State University Academic Aptitude Examination, The Pennsylvania State University, 1947.

AN ANALYSIS OF TEST SCORES AND GRADES FOR PREDICTING SUCCESS OF COLLEGE STUDENTS IN ENGLISH COMPOSITION

WILLIAM E. KUNHART, LIONEL R. OLSEN
Hartnell College, Salinas, California

The Problem

Educators frequently use high school grades and standardized test scores to provide predictions of student success in college course. In this study the writers desired to ascertain the effectiveness of grades and standardized test scores for predicting success of students in freshman college English. The study pertained to students who scored between the 30th and 75th percentiles on the Cooperative English Test (CET). Students scoring between these percentiles were required to pass English A, a review of high school English, with a C or better before their placement in college English composition. The group below the 30th percentile were required to take a remedial type course. Those students who scored at the 75th percentile and above were placed directly into college English composition, and presented no counseling-placement problem.

The Procedure

Two groups of students were studied in this research. In the first group (N = 163) three variable factors were used to determine effectiveness in predicting success in freshman composition. The variables were:

1. English A grade (an English class provided for intering deficient freshman students).
2. High school intelligence score (group test).

3. High school senior English grade.

The correlations with the English composition grade were as follows: English A grade +.303, IQ, score -.052, and high school English twelve grade +.306. A multiple correlation (R) of .418 was secured.

The second group in this research (N = 50) was taken from the original group of N = 163, which included information on the following variables:

1. English A grade.
2. High school intelligence score (group test).
3. High school English (grade twelve).
4. Cooperative English Test Total Score.
5. American Council of Education Psychological Test Linguistic Score.

The correlations in the English composition grade were as follows: English A grade +.333, IQ, score +.250, high school English grade twelve +.277, Cooperative English Test Total score +.307, and the American Council of Education Psychological Test Linguistic score +.174. The multiple R from this group was .531.

Conclusion

It would appear from the data that the predictions will be somewhat improved by securing additional information about grading and testing procedures. Even then, predictions should be made with extreme caution.

JOURNAL *of* EDUCATIONAL RESEARCH

Volume 52

November, 1959

Number 3

CONTENTS

A Comparison of Forced vs. Free Q-Sort Procedure—Robert D. Hess, Douglas L. Hink	83
Effect of Familiar Background Music Upon Film Learning—John Freeman, Charles O. Neidt.....	91
A Research in Mental Arithmetic Involving Subtraction—Herbert T. Olander, Betty Irene Brown	97
Sex Differences Among High-School Seniors—Pairlee J. Stinson, Mildred M. Morrison	103
Prediction of Success in Graduate School at Rutgers University—Milton M. Schwartz, F. Eugene Clark.....	109
Acceptance and Rejection as Related to Length of School Attendance—Jack Frymier.....	112
Some Attitudinal Differences Among Educational Specialists, Administrators and Teachers—Bartholomew D. Wall	115
Intellectual and Personality Characteristics of University of Utah Students—Frank B. Jex, Reed M. Merrill.....	118

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

A COMPARISON OF FORCED VS. FREE Q-SORT PROCEDURE*

ROBERT D. HESS and DOUGLAS L. HINK
University of Chicago

IN THE VARIOUS recent discussions concerning techniques for intra-individual assessment, Q-methodology has received more attention than any other single type of analysis. William Stephenson has been instrumental in bringing this method to the fore. His book on the subject gives an account of his views on the premises, promise and utilization of Q-methodology (6).

Among the reviews of Stephenson's book, Cronbach and Gleser's discussion seems the most definitive in both its positive and negative evaluations (1). One phase of Q-technique as advanced by Stephenson that these reviewers find particularly inappropriate is the use of a forced distribution of the items, i. e., requiring that a predetermined number of items be placed in each cell of the sort. They believe that this procedure negates the possibly advantageous use of analysis of variance. Furthermore, they point out that the F-test is based on the assumption that values in each cell are randomly sampled from an infinite population and that this assumption does not hold in the forced distribution. Cronbach and Gleser also object to the use of the near-normal distribution as discarding information about differences in scatter, although noting that it ensures variance in responses and eliminates response sets.

Stephenson's principal argument for the use of the forced Q-sort centers around the importance of requiring the ordering of subjective data into a semi-normal distribution. Stephenson contends that the means are of little significance while variations are extremely useful. Cronbach and Gleser's remarks that the forced distribution insures variance in response and eliminates response sets also argue favorably for this type of administration.

However, Stephenson appears agreeable to the use of non-forced distribution in suitable situations. The loss of differences in scatter as pointed out by Cronbach and Gleser may not be important in all research designs, yet it is a factor to consider in any use of the Q-sort. In relation to the treatment of data, their argument is to discard the forced-

sort since it negates the use of the analysis of variance which is potentially advantageous.

The Research Problem

The purpose of the research reported here was to evaluate free and forced Q-sort administrations on the basis of comparative merit in relation to the objectives of a particular research. The study was conducted in the context of a larger research on Erikson's concept of the process of identity formation in adolescence (3, 4). Identity for Erikson suggests "a social function of the ego which results, in adolescence, in a relative psycho-social equilibrium essential to the tasks of young adulthood" (4). The focus of the major project lay not in discovering the problems of adolescence but rather the way in which the person derives a stable concept of himself and his place as a member of society.

Traditional methods of exploration into areas of personality such as protective tests, interviews and questionnaires were administered as part of the research design. However, it was believed that a specialized instrument directly related to the major problem under consideration might facilitate and improve the assessment techniques. The Q-sort was selected as the instrument for these purposes. However, Cronbach and Gleser's criticisms of this instrument left several questions to be answered concerning the method of administration. Since the larger research plan called for intra-individual comparison over time, an evaluation of the relative usefulness of free versus forced-sort procedures was of practical as well as theoretical significance.

Research Design--The Identity Q-Sort

In order to make use of Stephenson's structured sample design, areas and dimensions applicable to the overall research inquiry were derived from Erikson's concept of identity. The three major areas of resolution selected from general theory and research on adolescence were 1) occupational

* Footnotes will appear at the end of the article.

choice, 2) sex role and 3) social role. Each of these areas subsumed dynamic factors also related to identity. These dimensions were: a) autonomy in making decisions and taking action free from domination of parents, significant adults or peers; b) feelings of competence in present or future role performance; and c) definiteness of the individual's identity decisions. The dimensions will be referred to as Autonomy, Competence and Definiteness. It should be noted that areas and dimensions may be combined, e. g., Occupational Autonomy, or may be discussed individually.

A series of four paired items was constructed for each combined area-dimension. The four items were paired into positive and negative statements for each combined area-dimension. This structure gave a total of 72 items. Ten intervals were used for these statements, ranging from cell one as "least like me" to cell ten as "most like me." The use of ten intervals has the advantage of requiring the test-taker to make at least a minimal decision on which side to throw the items since there is no cell at the mid-point. It was believed this should act against the tendency to use the median cell to avoid making decisions, particularly in the free-choice administration.

Procedure

First semester high school juniors were selected as the population for study. High school graduation with the attendant necessity for decisive action was distant enough to preclude social pressure as a dominant factor and yet this ceremony was near enough to be perceived as an event calling for personal decision. A metropolitan high school was chosen which had a relatively homogeneous student body drawn primarily from the "common-man" social classes with a relatively small group from upper-middle class families. The school administrators speak of their students as "good citizens", in the usual scholastic connotation of the phrase. They also note with justifiable pride that an unusual number (over sixty per cent) of their students enter college.

The sample for this study was selected from the class list by the use of random numbers. As a rough index of its randomness, the sample derived was in close correspondence to the sex ratio of the population. The Otis Aptitude Test showed a range of percentiles based on national norms from 13 to 96 with a median of 64.

The sample thus selected was divided to allow for the different types of administrations to be given in diversified order. An "intensive" sample and two "reliability" samples were obtained, divided equally as to sex, and the "intensive" sample was further segmented into a lower socio-economic group and a middle socio-economic group to permit an examination of possible socio-economic status differences in test behavior. The reliability samples were used to study the stability of the identity

Q-sort on a test-retest of the same type of administration within a time period of one week. The intensive sample was given both the free- and forced-sorts with half of the sample taking the free-sort first and 50 per cent taking the forced-sort first. This procedure resulted in the following sub-samples:

1) Free-choice reliability. Five males and five females were given a test-retest of the free-sort with an interval of one week between administrations.

2) Forced-choice reliability. Four males and six females were given a test-retest of the forced-sort with an interval of one week between administrations.

3) Intensive sample. Sixteen males and seventeen females: a) Nine males and eight females were given the free-sort, followed in one week by the forced-sort. b) Seven males and nine females were given the forced-sort, followed in one week by the free-sort.

The Q-sort was given in groups of four, with all members of a given group taking the same type of administration. The 72 items were printed on small cards and given to the student in a predetermined random order. Those taking the free-sort were allowed to arrange their cards on a table in any manner and distribution they desired within the range of ten sorting intervals. Each member of the forced-choice group was seated in front of an upright board on which the intervals were numbered from one to ten. The predetermined number of items to be placed in each cell was indicated by a like number of nails upon which to hang the cards². The forced distribution was as follows:

Interval	1	2	3	4	5	6	7	8	9	10
No. of Items	3	5	7	9	12	12	9	7	5	3
% of Items	4	7	10	12	17	17	12	10	7	4

Instructions for the two types of administration were similar except for the requirements of item distribution. Those taking the free-sort were informed that they could place as many or as few items in each cell as they desired while those taking the forced-sort were told to place only one item on each nail of the board. The method of ordering, the division of the cells into "most and least like me" sections and the mid-line between intervals five and six were demonstrated on the blackboard for those taking the free-sort. Both groups were individually queried at the end of the testing period to ascertain that they understood the instructions.

The positive and negative items in each area-dimension were separately summed, each item being given the numerical weight from one to ten of the cell into which it was tossed. To reduce the magnitude of these scores for the area-dimensions and to derive a component score by combining the positive and negative items of the area-dimensions, the negative items were subtracted from the positive items.

Results--Reliability Sample

Our interest centered around the areas and dimensions, rather than the individual items or the total score. Since the consistency of the ranking of the area-dimensions by individuals was more important to the major research project than absolute raw scores, the rho rank order coefficient was the most appropriate statistical device. It also required no assumption about the type of distribution (2). The rank order correlations that were obtained are shown in Table I.

The same rho rank order coefficient was used to test the stability of intra-individual ratings of the nine area-dimensions on the two types of administration. The results obtained are shown in Table II.

Of the differences or changes in ranking equal to or greater than three in the intra-individual rank order of the nine area-dimensions, the three lowest correlations contributed 73% of these differences to the forced "reliability" sample and 71% to the free. A drop from .69 to .33 in the forced reliability sample accentuates the division of correlations.

Intensive Sample

Pearson product moment correlations were run between free- and forced-administrations on areas and dimensions for the intensive sample of thirty-three, giving the results shown in Table III.

The correlations between free and forced administrations range from .60 to .76, the areas range from .60 to .77, and the dimensions range from .66 to .76.

Intra-individual rho correlations of the nine area-dimensions were also computed for the 33 members of the intensive sample. The coefficients resulting for the two groups who were given the two types of administration are shown in Table IV.

Distribution of Items in Cells

Results of the first administration of the free-sort to the reliability sample of 10 were combined with the data from 17 subjects in the intensive sample who took the free-sort first. This provided a distribution of the number of items in each of the ten cells for initial free-sort administration. The following distribution was obtained:

Interval	1	2	3	4	5	6	7	8	9	10
% of Items	15	10	8	7	9	9	8	9	10	15

These figures show that 30% of the items are thrown in the extreme intervals, i.e., 1 and 10, when the sample is given complete freedom as to the number of items that may be placed in each cell.

The free-sort reliability sample showed the following distribution of percentages on the second administration:

Interval	1	2	3	4	5	6	7	8	9	10
% of Items	14	8	7	7	10	13	12	8	8	13

Effects of response set introduced by instructions on the forced-sort may be observed in the subsequent free sorting behavior of the forced-free intensive group. The distribution of items by interval is as follows:

Interval	1	2	3	4	5	6	7	8	9	10
% of Items	8	11	10	9	10	12	10	11	10	9

Discussion

In the rank order correlations of the area-dimensions, coefficients not significant at the 5% level were found in two area-dimensions, one total area, and one total dimension in the forced-choice reliability group. Four area-dimensions and one total dimension were independent in the free-choice test-retest group. The one total dimension (Competence) was the only area-dimension which proved unstable in both types of administration. Since the uncontrolled variable was the type of administration and, with one exception, different area-dimensions were unstable in each group, it would appear that either the two types of administration are measuring different things, or the unevenness of area-dimension stability is due to individual differences. Often in psychometrics the first would be assumed to be the cause of the incongruity. However, since the items, length of time between administrations, and number of cells were constant, it seems more probable that the influence of intra-individual variations is responsible for the fluctuations of area-dimension stability between the two types of administration.

The intra-individual rank order correlations also indicate that individual stability is an important factor for consideration. This is suggested in the fact that in each reliability sample, three individuals (30%) showed insignificant coefficients, on a test-retest in one week under similar conditions of administration. The large proportion of gross changes in ranking of area-dimensions in the test-retest behavior of these six individuals provides support for the contention that instability of area-dimensions arose from individual variations rather than instability of the area-dimensions per se. The large percentage of gross changes indicates these are highly variable individuals.

The high percentage of males with insignificant intra-individual rank correlations coefficients in both the forced-free intensive sample and the reliability samples is another significant factor. Males in this age group are considered to be less mature in all areas than females so that test-retest stability may prove to be one criterion of maturity level in this group.

From the proceeding discussion of the results of this research, it appears that the two types of administration gave similar results when the identity Q-sort was used with adolescents. Low area-dimension correlations may be attributed largely to intra-individual variation in a test-retest within one week. This individual variation may in turn be related to adolescent instability in the search for

TABLE I

TEST-RETEST CORRELATIONS BY AREA-DIMENSION FOR FORCED AND FREE
CONDITIONS OF ADMINISTRATION (RELIABILITY SAMPLE)

Area-Dimension	Forced Test-Retest ³	Free Test-Retest ³
Occupational Autonomy	.88	.32*
Occupational Competence	.76	.52*
Occupational Definiteness	.62	.84
Social Role Autonomy	-.12*	.80
Social Role Competence	.72	.50*
Social Role Definiteness	.58	.26*
Sex Role Autonomy	.68	.59
Sex Role Competence	-.18*	.89
Sex Role Definiteness	.83	.87
Total Occupation	.82	.92
Total Social Role	.41*	.65
Total Sex Role	.78	.57
Total Autonomy	.65	.72
Total Competence	.15*	.44*
Total Definiteness	.83	.88
Identity Total	.79	.80

TABLE II
INTRA-INDIVIDUAL TEST-RETEST RANK ORDER CORRELATIONS
(RELIABILITY SAMPLE)

Forced Test-Retest ⁴	Sex	Free Test-Retest ⁴	Sex
.99	F	.95	F
.98	M	.90	F
.90	F	.89	M
.88	F	.86	M
.88	F	.75	F
.75	F	.72	M
.69	M	.71	F
.33*	M	.52*	F
.14*	M	.43*	M
.08*	F	.28*	M

TABLE III
CORRELATIONS BETWEEN RESPONSES ON FREE AND FORCED ADMINISTRATION
FOR AREAS AND DIMENSIONS (INTENSIVE SAMPLE)

Occupational Autonomy	.64	Total Occupation	.60
Occupational Competence	.60	Total Social Role	.66
Occupational Definiteness	.60	Total Sex Role	.77
Social Autonomy	.62	Total Autonomy	.69
Social Competence	.64	Total Competence	.66
Social Definiteness	.76	Total Definiteness	.76
Sex Role Autonomy	.72		
Sex Role Competence	.66		
Sex Role Definiteness	.64	Identity Total	.75

TABLE IV

INTRA-INDIVIDUAL CORRELATIONS BETWEEN RESPONSES ON
FREE AND FORCED ADMINISTRATION (INTENSIVE SAMPLE)

Free-Forced	Sex	Forced-Free	Sex
1.00	F	.98	F
.96	F	.90	F
.90	F	.88	F
.90	M	.87	F
.85	M	.80	F
.82	M	.78	F
.78	M	.77	M
.71	M	.75	M
.69	F	.65	F
.62	F	.63	F
.58*	F	.62	F
.46*	M	.57*	M
.42*	M	.30*	M
.42*	M	.09*	M
.39*	F	.02*	M
.25*	M	.00*	M
.15*	F		

identity.

Although self-concept is the common terminology used in connection with Q-sorts of this type, we suggest that self-representation may be a more accurate term. In this context, self-representation would be the manner in which the individual presents himself to the outer world, whereas self-concept is his private image of himself. Thus, the Q-sort would be self-representation which may or may not diverge from the person's self-concept. Instability of self-representation over a brief span of time may be a result of self-concept instability or may be an inability or reluctance to portray this accurately to others. A stable self-representation may be indicative of a stable self-concept or reflect a facade behind which lies an unstable or unacceptable self-concept.

Three major causal factors related to the degree of stability in adolescent self-representation may be hypothesized. First is the problem of anxiety and defense in communication. In the identity Q-sort we requested the individual to rate himself in various areas and dimensions. If anxiety is pronounced in any of these areas, it is likely that poor defenses there will cause instability of self-representation. Secondly, the individual's concern with his appearance in the eyes of others, as compared with what he feels he is, will affect stability. That is, the individual's ability to relate self-concept and self-representation may manifest itself in the degree of stability. Third is the relation of stability of self-representation and role-portrayal. If the individual's role is not clearly defined in these areas, self-representation instability will result. All of these hypotheses are particularly relevant to a study of the adolescent's search for identity.

The question of relative appropriateness of free- or forced-administration has not been fully answered by our research results. The lack of clear statistical superiority of either technique for our purposes and the psychological significance of responses obtained on both procedures suggested to us a method which combines the advantages of both. The forced distribution used in this study requires that one-third of the statements be thrown in the two center piles and only one-twelfth of the cards in intervals one and ten. This reduces the possibility of wide variation in the area-dimensions since each of these is covered by only eight statements. The free-choice has an equally disagreeable factor in that it allows for extreme acceptance or rejection of statements by throwing a large percentage in the intervals on either end of the continuum. This again allows for a more stable self-representation on test-retest than may actually be present.

In order to incorporate the best of each type of administration, a new type of administration seems possible. For our identity sort, this is a procedure that would require the subject to throw at least six but no more than ten statements in each of the ten intervals. The forcing would require a discrimination along the continuum, yet the 12 free-choice

items would give important information on scatter and response idiosyncrasy and make for less feeling of restriction on the part of the subject.

Livson and Nichols have recommended a forced, rectangular distribution of items to obtain maximum reliable information by requiring a maximum possible number of discriminations (5). They also note that Block, in an article which is in press, found that "variance in elevation and scatter reflects intersorter response idiosyncrasy, but makes little, if any, contribution to discrimination among objects." (5, p. 160)

However, in both of these research designs the investigators were interested in reliable descriptions of social objects rather than personal description as in the present study. By combining the forced-rectangular and free types of administration, important information should be gained in idiosyncratic material while providing reliable information. Both types of data are important in the study of stabilization in and between self-concept and self-representation.

One of the primary advantages of the use of intra-individual stability in studies involving Q-methodology is that it aids in preventing biasing of results by external factors. In the majority of research studies involving Q, there is a tendency to judge differences between Q-sort responses under different sets of instructions and over long periods of time as valid on the assumption that an administration is a "true" representation of the individual's self-concept. If self-sorts of some individuals vary widely within one week, differences over a longer period of time may be related as much to the time interval per se as to experimental variables introduced.

In testing for stability, externally determined factors are decreased. The individual is requested to give a self-sort, unaware that he will be asked to take the same sort again within a brief span of time. The incongruity between these two sorts is therefore an indication of instability in his self-representation. A shorter span of time might prove an even better indicator of individual stability. As noted earlier, whether this instability is in agreement with self-concept or merely an indication of an unsuccessful attempt to distort the self-representation must be discovered through other instruments such as interviews and projectives.

Summary

A Q-sort was constructed to explore Erikson's concept of identity formation in adolescence. A balanced design was constructed covering the areas of occupation, social role and sex role and the dimensions of autonomy, competence and definiteness. A total of 72 items was paired, one item dealing with "positive" identity while its mate dealt with the "diffuse" identity counterpart. This gave 12 pairs of items in each dimension, or four pairs of items in each of the nine combined area-dimensions.

A ten interval scale was used for the distribution of the items.

The sort was administered in groups of four to juniors in a metropolitan high school by free- and forced-choice methods. Those taking the free-sort administration were allowed to place as many or as few Q-sort items in each cell as they desired.

As a result of correlations (ρ) between areas, dimensions, and area-dimensions there appeared to be no definitive answer as to which sort was most reliable for our purposes. An intra-individual rank correlation found three in each reliability sample whose ranking of the nine area-dimensions was below the critical value at the 5% level of significance. These cases contributed 73% of the intra-individual rank order differences equal to or greater than 3 (in the nine area-dimensions) for the forced group and 71% of such differences in the free test-retest group. These results indicate that insignificant correlations among area-dimensions are largely a result of extreme intra-individual variation.

The Pearson product moment correlation between free- and forced-choice in the intensive sample showed correlations in the areas, dimensions and area dimensions ranging from .77 to .60. These correlations show that the free and forced sorts do not give strikingly different results.

FOOTNOTES

- *1. This project was supported by a grant from the Social Science Research Committee of The University of Chicago.
2. We are indebted to Mr. John Shlien of the Counseling Center, University of Chicago, for

suggesting this procedure.

3. Correlation coefficients followed by an asterisk are those which are below .564, the figure given by Dixon and Massey as the one-sided critical value of the rank correlation coefficient at the 5% level when N is ten (2, p. 261).
4. The intra-individual correlation coefficients followed by an asterisk are those below .600 which Dixon and Massey list as the one-sided critical value for a rank correlation coefficient at the five per cent level of significance when N is nine (2, p. 261).

REFERENCES

1. Cronbach, Lee J., and Gleser, Goldine C., "Book Review: William Stephenson, *The Study of Behavior*," *Psychometrika*, 19 (1954), 327-331.
2. Dixon, Wilfrid J., and Massey, Frank J., *Introduction to Statistical Analysis*. New York: McGraw-Hill Book Co., 1951.
3. Erikson, Erik H., *Childhood and Society*. New York: Norton, 1950.
4. Erikson, Erik H., "The Problem of Ego-Identity," *Journal of the American Psychoanalytic Association*, 4 (1956), 56-121.
5. Livson, N. H. and Nichols, T. F., "Discrimination and Reliability in Q-Sort Personality Descriptions," *Journal of Abnormal and Social Psychology*, 52 (1956), 159-165.
6. Stephenson, William, *The Study of Behavior: Q-Technique and Its Methodology*. Chicago: The University of Chicago Press, 1953.

EFFECT OF FAMILIAR BACKGROUND MUSIC UPON FILM LEARNING

JOHN FREEMAN and CHARLES O. NEIDT
University of Nebraska

A FREQUENTLY expressed point of view regarding the selection of background music for educational films is that such music should be unfamiliar to the audience so that attention will not be drawn from the film content. Lindgren (5, p. 147) expressed this point of view when he said:

"Music in a film may call attention to itself not only where it is too good, but also where it has a strong and clearly defined melodic line. The spectator's attention is drawn to the melody and the composer who wishes to keep his music in the background will work rather on changes of harmony and rhythm. The use of well-known music is even more distracting, and has the additional disadvantage that it often has certain associations for the spectator which may conflict entirely with the associations the producer wishes to establish in his film. The use of classical music for soundfilms is entirely to be deplored."

It is conceivable that there are many university film producers and many smaller commercial producers who would use film music if there were more latitude in the choice of compositions. Experimental evidence regarding the use of familiar or classical music for film backgrounds is scant, and what little there is is inconclusive (4, 9).

The purpose of the present study was to determine the effect upon film learning of using familiar compositions as background music as contrasted with unfamiliar music. Stated otherwise, this study was conducted to test the hypothesis that familiar music distracts from the visual and verbal elements of the film. To accomplish this purpose, two experimental versions of a factual film were shown two groups of subjects, and the subjects were tested over the information contained in the film.

The Experiment

Selection of the Film: A ten-minute color film entitled "Color Categorizing Behavior of Rhesus Monkeys" (1) was selected for the experiment. This film was chosen because of the amount of informa-

tional material which it contained, and because the setting of the film was the same throughout. Thus, distracting visual material was avoided.

Selection of the Music: The unfamiliar music was selected by the editor¹ of a commercial film music company from their library of recorded film music. Four compositions were used in the unfamiliar film version, "Hurly Burly," "The Laughing Cavalier," "Playground," and "Joie de Vivre." In editing the music for the film, nine excerpts from the four compositions were used. Musical breaks were timed to coincide with natural transitions of the film.

Choice of familiar music was purely arbitrary, although music was selected which closely resembled, in mood and tempo, the music used for the unfamiliar music track. Disregarding the familiarity factor, this music was considered to be the type which would be used by film producers.

The following compositions were chosen: "Fiddle Faddle", "Hora Stacatto", "Humoresque", "Flight of the Bumble Bee", "Anitra's Dance", "Il Bacio", three pieces from the "Nutcracker Suite", and "Sabre Dance". Originally, "Dance of the Tumblers" from Rimsky-Korsakov's "Snow Maiden" was used in the familiar music track, but a pretest showed that this composition was not familiar, and it was deleted. The familiar music was edited so that the breaks occurred at the same points in the film as in the unfamiliar music version.

Preparation of the Final Tracks: Two ten-minute tapes containing both voice and music were prepared by using three recorders, one playing the voice tape, one playing the music tape and the third recording both voice and music. During the dubbing, a monitor system allowed fading the music in and out--low when the narrator was speaking, and high when the narrator was not speaking. The same voice track was used for both the familiar and unfamiliar sound tracks. The final result was two ten-minute sound tracks, differing only in the type of music used in the background.

1. Mr. Gene Moore, The Calvin Company, Kansas City, Missouri.

The Test Battery: To determine the extent of film learning, a test covering the informational material of the film was constructed. On the basis of the results from a pretest, twenty-five items were selected for the final edition of the film test which was to constitute the criterion for the experiment. The time allowed for responding to these twenty-five items was fifteen minutes.

Because two independent groups of subjects were to be used in the experiment, it was felt desirable to control other variables presumably related to the criterion. The control variables selected for this investigation were scores on a music identification test, a music familiarity test, and a vocabulary test. The vocabulary test was used to control individual differences in learning aptitude between the two groups.

To obtain measures of the control variables, music identification and music familiarity, a music literacy test was constructed. This test consisted of twenty-five short excerpts from musical compositions, including the four unfamiliar pieces and the seven familiar pieces used in the two versions of the film. The test was constructed so that the subjects would make two responses to each musical excerpt. First, they would identify the music by name or by composer from a four-choice list of responses for each excerpt. Second, they would designate the music's familiarity as "Very Familiar", "Somewhat Familiar", or "Unfamiliar". These responses would be made as a tape containing the music was played. The time required for both sections of the music literacy test was ten minutes.

The vocabulary test used to measure verbal ability was a sixty-item unpublished test which had been constructed by the experimenter as a subtest for an experimental edition of a graduate aptitude test. The items in this test were of the same-opposite type, and the time allowed for responding to the sixty items was five minutes.

The Subjects: The subjects in the experiment represented six undergraduate and two graduate classes at the University of Nebraska. Group I, (93 subjects) which saw the familiar music version of the film, consisted of two undergraduate educational psychology classes, one undergraduate sociology class, and one graduate educational psychology class. Group II, (94 subjects) which saw the unfamiliar music version of the film, consisted of the same number and type of classes. Members of the eight classes were randomly assigned to the two groups.

Experimental Procedure: Throughout the experiment, a careful control on physical conditions was attempted. The same room was used for each of the showings. Volume of sound for the film and for the music literacy test was held constant. Verbal instructions for each group were the same. The same projector, tape recorder and speaker were used for the showings.

When a group had assembled, they were given a booklet containing the test materials used in the in-

vestigation. As soon as the group had completed the informational material on the front of the booklet, the subjects were told that they were participating in a film learning experiment, that they would see a ten-minute film, and that they would be given fifteen minutes to respond to the items in the film test. When the film had been shown and the subjects had completed the film test, the music literacy test was administered. The music literacy test was followed by the five-minute vocabulary test.

Results

In Table I are shown the mean scores obtained by the two groups of subjects on the criterion variable and the three control variables used in the investigation. The film test, the music identification test, and the vocabulary test were scored one and zero for correct and incorrect responses. The music familiarity test was scored two for a "Very Familiar" response, one for a "Somewhat Familiar" response, and zero for an "Unfamiliar" response. Inspection of the data shown in Table I reveals that the unfamiliar music group obtained slightly higher mean scores on the film test, the music identification test and the music familiarity test than the familiar music group. The familiar music group, however, obtained the higher mean score on the vocabulary test.

To determine the significance of the difference between the criterion means of the two groups with scores on the two sections of the music literacy test and score on the vocabulary test held constant, the analysis of covariance was used. This statistical technique involves testing the significance of the difference between the criterion means after they have been adjusted for individual differences among the control variables. Although accomplishing the same result as pairing cases on the basis of the control variables, the analysis of covariance has the advantage of permitting use of all data and eliminating the laborious pairing procedure.

The results of the application of the analysis of covariance to the data obtained from this investigation are shown in Table II. The F-value of .14 with 1 and 182 degrees of freedom is far from significant. The hypothesis, "there is no significant difference between the two film test means with scores on the music identification test, music familiarity test, and vocabulary test controlled" was therefore not rejected.

In the use of the analysis of covariance, the assumption is made that the control variables are significantly related to the criterion. To the extent to which the control variables are unrelated to the criterion, the process of statistical control is ineffective. To test the effectiveness of the controls used in this investigation, an analysis of regression was made and is shown in Table III. Since the F-value of 10.26 with 3 and 183 degrees of freedom is considerably larger than that required at the 1 per cent level, it was concluded that the multiple cor-

TABLE I
MEAN SCORES FOR THE CRITERION AND CONTROL OF VARIABLES

Variable	Familiar	Mean Score Unfamiliar	Total
	N = 93	N = 94	N = 187
Film Test (Criterion)	14.48	14.60	14.54
Music Identification	10.04	10.46	10.25
Music Familiarity	21.95	23.93	22.94
Vocabulary	28.08	26.72	27.40

TABLE II
SIGNIFICANCE OF THE DIFFERENCE BETWEEN FILM TEST
MEANS WITH CONTROL (ANALYSIS OF COVARIANCE)

Source of Variation	Degrees of Freedom	Residuals		F
		Sum of Squares	Mean Square	
Total	183	1,639.54		
Within Subgroups	182	1,638.24	9.0	.14
Difference	1	1.30	1.30	

TABLE III

ANALYSIS OF REGRESSION OF THE FILM TEST SCORES AND THE CONTROL VARIABLES

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F
3-variable regression	3	273.92	93.91	10.26
3-variable residuals	183	1,628.53	8.90	
Total	186	1,902.45		
$R = \sqrt{.143893} = .379$				

TABLE IV

MEAN PERCENTAGE OF RESPONSES TO DIFFERENT TYPES OF MUSIC
IN THE MUSIC LITERACY TEST (187 SUBJECTS)

Type of Music	Number of Selections	Correct Identification	Very Familiar	Somewhat Familiar	Unfamiliar
Unfamiliar	4	5	1	12	87
Familiar	7	71	56	30	14
Other Music	14	36	26	38	36

relation coefficient, $R_{Y(X_1X_2X_3)}$ of .379, is significantly different from zero. Thus the rejection of the hypothesis of zero correlation provides evidence that the controls were effective in eliminating individual differences in the characteristics used as controls.

Analysis of the Tests: An item analysis of the music literacy test confirmed the judgment of the music selectors in designating the music as familiar and unfamiliar. It will be seen from Table IV that the mean percentage of the 187 subjects correctly identifying the four unfamiliar selections was low. In addition, the mean percentage of subjects marking these four unfamiliar selections as "Very Familiar" was 1 per cent, whereas the mean percentage of subjects marking them as "Unfamiliar" was 87 per cent.

For the seven familiar compositions the mean percentage of correct identification was 71 per cent. Ninety-five per cent of the subjects correctly identified "Flight of the Bumble Bee", and 94 per cent identified "Sabre Dance". The mean percentage marking these seven familiar compositions as "Unfamiliar" was 14 per cent.

For the other fourteen compositions in the music literacy test, percentages of correct identification and indications of familiarity fell between those of the familiar and unfamiliar selections.

Reliability coefficients were determined for the film test, the music identification test, the music familiarity test and the vocabulary test. The Spearman-Brown estimates of the reliability coefficients for these tests were as follows: film test, 0.63, music identification test, 0.71, music familiarity test, 0.73, and vocabulary test, 0.84. It will be recalled that the film test consisted of twenty-five items and required fifteen minutes for administration, the music literacy test, twenty-five items and ten minutes, and the vocabulary test, sixty items and five minutes. Although these reliability coefficients are not extremely high, the four tests were considered sufficiently reliable for purposes of this study.

Discussion

The compositions for the familiar music version of the film were chosen to represent the type of music which is adaptable to motion pictures. Popular tunes were not considered. Since selections were made from classical or semi-classical music, it was not expected that subjects would achieve one hundred per cent identification for these compositions. It is conceivable that larger differences between the mean film test scores of the two groups would have resulted if the audiences had been more familiar with the background music chosen for the familiar music version of the film. Further research to determine the degree of music familiarity which will result in significant distraction is suggested.

Another possible explanation of the results of this study is that the subjects in both groups were relatively unaware of the musical background of the film. Several subjects, after the showing, indicated that they could not be sure that music accompanied the film throughout. This finding parallels that of Nuckols and Abramson (6) who investigated the effect of two versions of the film "We Make Fire". One version had a continuous musical accompaniment and the other had music only during introductory and final titles. Pupils in the third, fourth and fifth grades were used as subjects. No differences were found in informational learning or in expressed attitudes resulting from the two versions of the film. More than half of the pupils who saw the "music" version failed to recognize afterward that music had accompanied the film throughout. Such evidence suggests the desirability of determining the relationship between familiarity of background music and audience awareness.

Generalizations drawn from the present study are subject to several limitations. The film used in the investigation was only ten minutes in length. The sample consisted only of university students. No attempt was made to evaluate learning outcomes other than the acquisition of information.

Summary

This study was undertaken to determine the effect of using familiar background music in a factual film. Two experimental versions of the film, "Color Categorizing Behavior of Rhesus Monkeys," were prepared and presented to two experimental groups. One group of ninety-three subjects saw the film with familiar background music. The other group of ninety-four subjects saw the film with unfamiliar background music. Immediately after the showings, both groups were given a twenty-five item multiple-choice test covering the informational content of the film. Both groups also took a music literacy and a vocabulary test, the results of which were used as control variables. When the analysis of covariance was applied to the data obtained from this investigation, no significant difference was found between the film test means of the two groups.

REFERENCES

1. Color Categorizing Behavior of Rhesus Monkeys, (Motion Picture) Coronet, Chicago, 1943. 10 Minutes, Sound, Color, 16 mm.
2. Feldman, J. and Feldman, H., Dynamics of the Film, Hermitage House, Inc., New York, N. Y., 1952.
3. Gibson, H. C., Films in Business and Industry, McGraw Hill Book Co., Inc., New York, N. Y. 1947.
4. Hoban, C. F., Jr. and Van Ormer, E., Instructional Film Research, 1918-1950, Office of Naval Research, SCD 269-7-19.

5. Lindgren, E., The Art of the Film, George Allen and Unwin, Ltd., London, 1948.
6. Nuckols, R. C., and Abramson, R., Effect of Musical Backgrounds in Instructional Films, unpublished manuscript, Project 20-E-4-7 Instructional Films Research Program, Pennsylvania State University, State College, Pennsylvania, 1949.
7. Rotha, P., Documentary Film, Faber and Faber, Ltd., London, 1952.
8. Strauss, L. H. and Kidd, J. R., Look, Listen and Learn, Associated Press, New York, N. Y. 1948.
9. Zuckerman, J. V., "Music in Motion Pictures," Review of Literature with Implications for Instructional Films, SDC 269-7-2, Special Devices Center, Port Washington, L. I., N. Y., 1949.

A RESEARCH IN MENTAL ARITHMETIC INVOLVING SUBTRACTION

HERBERT T. OLANDER, University of Pittsburgh
BETTY IRENE BROWN, Albion, Pennsylvania

Introduction

The term "mental arithmetic" as here used applies equally well to other similar terms, such as "oral arithmetic" and "computation without pencil and paper." In a sense, of course, all computation is mental.

Mental arithmetic was very popular in the early history of our schools, at which time it was justified on the basis of faculty psychology or the theory of formal discipline. When the latter purpose for inclusion in the curriculum fell into disrepute, mental arithmetic all but disappeared from arithmetic textbooks. Lately, it has again begun to receive some attention in the arithmetic program, now under a new justification—its value in everyday living. The renewed interest in this subject is also evident in certain recent researches such as those of Boulware (1), Hall (3), Petty (4) and Flournoy (2).

Problem

The research in mental arithmetic in the area of subtraction here presented deals with four questions:

1) To what extent does proficiency in mental arithmetic depend upon whether the exercises are presented orally to pupils or by means of flash cards? How does such proficiency compare with that possible when pupils receive the exercises in written form and can use paper and pencils in their calculations?

2) To what degree is proficiency in mental arithmetic, when the exercises are presented orally, dependent upon the factors of school grade, sex, general intelligence, general arithmetic achievement, and memory span?

3) How do thought processes and methods of solution involving mental arithmetic differ among high

and low achieving pupils?

4) How do such thought processes and methods vary with school grade, sex, intelligence, general arithmetic achievement, and memory span level of pupils?

Procedure

In all, more than 1700 pupils in Grades 6 through 12 from one particular school system were involved in this study. Approximately 300 of these pupils from Grades 7, 9, and 12 were used in determining the reliability of three forms of a 26-item subtraction test. The remaining 1400 pupils, approximately 200 in each grade from Grade 6 through 12, were used in the main study. These pupils were given one form of the test orally, another through the use of flash cards, and the third in written form. In the case of the first form of the test, the pupils were given the subtraction exercises orally and required to place their answers, calculated mentally, on answer sheets. The exercises in the second form of the test were presented by means of flash cards, the pupils again calculating without paper and pencil, placing answers on answer sheets. The third form was presented as a written test, the pupils being able to use paper and pencil in their solutions. The subtraction test consisted of 14 exercises involving whole numbers, three involving decimals, three with fractions, three with denominate numbers, and three involving verbal problems.

The five per cent of the pupils from the seven grades who scored highest on the subtraction test which was presented orally were interviewed so that their methods of work could be studied. Also, the five per cent scoring lowest were similarly interviewed. The high and low groups were quite evenly distributed through the seven grades. Seventy-two pupils composed the high achievers and 71 the

* For a complete description of this research, see A Study in Mental Arithmetic: Proficiency and Thought Processes of Pupils Solving Subtraction Examples, by Betty Irene Brown, (Unpublished Ph.D. dissertation, University of Pittsburgh, 1957).

low. During the interviews each of the 143 pupils was asked to work a minimum of 12 examples orally. Four to six additional examples were worked with paper and pencil so that a comparison could be made with the work designated as mental arithmetic. Interview data were recorded not only on special sheets designed for this purpose but on tape as well. The pupils were unaware of the tape recordings.

In addition to the subtraction test presented in the three forms—orally, with flash cards, and written,—a specially constructed memory span test was given. This was adapted from the Stanford-Binet Intelligence Scale, consisting of six levels of difficulty. Pupils were required to write on answer sheets four to eight digits either as presented or in reverse order, depending upon the level of difficulty of the item.

Intelligence tests had been given to all pupils involved in the study. Results from the Stanford Achievement test were available only for pupils in Grades 6 through 9.

Findings

A. Proficiency in Mental Arithmetic in the Area of Subtraction in Relation to the Method of Presentation.

1. Based on central tendencies of scores--

Pupils from Grade 6 through 12 scored on the average 50.4 per cent on the test items when presented orally and 78.8 per cent when presented by means of flash cards. An average score of 88.9 was achieved when pupils were given the test in written form and could compute through the use of paper and pencil. Hence, seeing the exercises through flash cards improved the proficiency in mental arithmetic by more than 28 per cent. A further ten per cent improvement was made possible when some of the so-called mental arithmetic could be done on paper. These differences in the per cents show that not only do pupils need practice in working with arithmetic exercises without paper and pencil but their greatest difficulty seems to lie in the task of visualizing numbers presented orally. In ordinary life situations, an individual finds considerable need for this ability.

2. Based upon correlations of test scores--

The correlations between performance on the test presented orally and through flash cards was .61. The correlation between proficiency in mental arithmetic when the exercises were presented orally and carrying on the same calculations by means of paper and pencil was .51. The latter correlation is the lower one probably because a pupil might be proficient in working examples with paper and pencil but less so when required to do the calculations "in his

head."

B. Proficiency in Mental Arithmetic Involving Subtraction, in Relation to School Grade, Sex, Intelligence, Ability in General Arithmetic, and Scores of Pupils on a Memory Span Test.

1. In relation to school grade--

Greatest improvement in mental arithmetic took place between the 6th and 7th grades and the 7th and 8th grades. Whether this reflected a developmental phenomenon or could be attributed to the program in arithmetic was impossible to determine.

2. In relation to sex--

Boys excelled girls in mental arithmetic. This was most evident when the arithmetic examples were presented orally rather than through flash cards. Girls showed superiority when paper and pencils could be used, though the difference was not statistically significant.

3. In relation to general intelligence--

No closer relation seemed to exist between intelligence and proficiency in mental arithmetic than between intelligence and ability to calculate with paper and pencil. In general this correlation was .50.

4. In relation to general arithmetic achievement--

Based upon results in only Grades 6 through 9, the correlation between ability in general arithmetic and mental arithmetic involving oral presentation was .65. Compared with the correlation of .50 between intelligence and mental arithmetic which involved a spread of eight rather than four grades, this is a high correlation, apparently indicating mental arithmetic is more dependent upon general arithmetic ability than it is upon intelligence.

5. In relation to memory span--

Before this study began, it was assumed that ability in mental arithmetic was dependent to a considerable extent upon a person's memory span. However, this expectation was not borne out by the results. The correlation between proficiency in mental arithmetic involving examples presented orally and memory span was found to be only .35.

C. Thought processes and Methods of Solution, Involving Subtraction, of high as Compared with Low Achievers.

More than 90 per cent of the combined group of high and low achievers used the "take-away" method of thinking subtraction and the "borrowing" procedure when one or more digits in the subtrahend exceeded the corresponding digits in the minuend. Thus a very few pupils used the additive method of thinking, the "carrying" procedure, or the "complementary" method (making use of 10). A few more of the high than low achievers used these latter methods.

Together with the methods mentioned in the paragraph above, six additional modes of thinking out solutions were found in use. These have been given the terms indicated below:

1. Counting Approach--

Example: $29 - 14 \dots$ "15, 16, 17; 18, 19, 20; that's 6; then 9 are 15." This pupil counted by groups of three ones. Counting tended to proceed by ones or tens. An example of the latter is: $71 - 44 \dots$ A high achiever said "54, 64, 74, minus 3, the answer is 27 because there are 3 tens minus 3." Counting required too much time and many low achievers forgot the amount counted before the example was completed.

2. Right-left Approach--

Pupil begins subtracting at right of figures, moving to left, from ones to tens to hundreds, etc. This is the usual method taught in schools. This method was generally used when pupils were permitted to use paper and pencils in solution of examples. Thirty-seven low and 17 high achievers used this approach consistently for all of their work.

3. Left-to-right Approach--

Example: 672
 -458

A pupil using decomposition or "borrowing" said "4 from 6, 2; 5 from 6, 1; 8 from 12, 4." Fifty-two high and 16 low achievers used this method. This sequence of thinking seems to have the advantage given by Gestaltists for their point of view. The parts are seen in relation to the whole of which they are composed. The analytic attack, or moving away from the whole, has a psychological advantage over the synthetic in which the thinking is toward the whole. In the latter type of thinking, the whole is often not sensed by the learner.

4. Rounding Approach--

This is based upon the importance of 10 in our number system. Rounding involved any multiple of 10, 5, or any combination of 10 and 5. Two variations of the use of 10 were observed: 1) round to a multiple of 10, e.g. in $51 - 34 \dots$ 34, 40, 6; 40, 51, 11; 17," and 2) round by a multiple of 10, e.g., in $51 - 34 \dots$ "34 and 10, 44; 44 and 7, 51; 17." Rounding by 10 seemed to be the easier of the two processes for the students. An advantage in rounding upward or downward was that the subtrahend or minuend, as the case may be, could be forgotten after the first step. Twenty-four high and 6 low achievers used this rounding procedure when solving mentally the subtraction examples presented to them orally by the observer.

a. Rounding Downward--

Example: $35 - 8 \dots$ "35 minus 5, 30; 30 minus 3, 27." This was usually done either by beginning with the exact min-

uend, e. g., in $35 - 8 \dots$ "35 minus 5, 30; 30 minus 3, 27," or by beginning with an approximation of the minuend, e. g., in $51 - 34 \dots$ 50 minus 30, 20; 11 minus 4, 7; 17."

b. Rounding Upward--

Example: $29 - 14 \dots$ "14 plus 10 plus 5; 15." Sometimes this was done very much like the procedure used by clerks in stores in handing back change to customers, e. g., in $51 - 34 \dots$ "34 plus 1, 35; 35 plus 15, 50; 1, 51; 17."

c. Compensatory Rounding--

Example: $51 - 34 \dots$ "50 minus 30, 20; minus 4, 16; plus 1, 17." In the example $34 - 9$, a pupil thought "34 minus 10 plus 1; 25." He explained that he thought "34 take away 10 are 24, but 10 is 1 too many to take away; therefore, the answer would be 1 more than 24, or 25."

5. Place-grouping Approach--

Example: $6000 - 2249 \dots$ "100 minus 49, 51; 60 minus 22, 38; minus 1, 37; 3751." In this approach, a pupil divided larger numbers into smaller ones composing them. For example, in $439 - 327$, when a pupil thought "39 minus 27, 12; 4 minus 3, 1; 112," he broke up the 439 into 400 and 39, and 327 into 300 and 27. Only five high achievers used this approach.

6. Number-relations Approach--

Example: $29 - 14 \dots$ "two 14's plus 1, 29; one 14 plus 1, 15." This method usually utilized multiplication or division as illustrated in this solution for $51 - 34 \dots$ "three 17's minus two 17's; 17." Only three pupils, high achievers, used this procedure.

The number of instances the 72 high and 71 low achievers used the six approaches to subtraction during the interviews, mentioned and illustrated above, is given below.

Approach	High Achievers	Low Achievers	Total
Counting	2	16	18
Right-to-left	67	71	138
Left-to-right	52	16	68
Rounding	24	6	30
Place-grouping	5	0	5
Number-relations	3	0	3

It will be noted that 1) more low achievers used counting, 2) more high achievers used the left-to-right approach and also the rounding approach, and 3) no low achiever used what has been termed the place-grouping and number-relations approaches.

Summarizing this part of the study, the following are some of the easily observed differences in procedures between high and low achievers:

1) High achievers had a greater tendency to use a variety of methods. Some used as many as four

of the six approaches listed in the table above. The difficulty of a subtraction example, such as the necessity for "borrowing", usually determined the method of solution.

2) The last four approaches listed in the table are especially suited to mental arithmetic (solving examples without the use of paper and pencil). Four times as many instances of the use of these four methods appeared among the high as among the low achievers.

3) Many low achievers but few high ones used counting as an approach.

4) Eighteen high achievers but only one low one often began solutions to examples the instant the first figure of the subtrahend became known. This practice no doubt was encouraged by the fact that each example was presented by the observer only once slowly and not repeated. The remaining pupils almost without exception said they repeated the whole example silently in order to remember it.

5) There seemed to be no advantage in the direction the pupils worked, i.e., downward from the minuend to the subtrahend or the reverse. About the same number of high as low achievers used each of the two procedures, though twice as many worked upward as downward.

The following comments volunteered by the high and low achievers are thought-provoking:

By High Achievers:

"I can't see it, I just remember it."

"I see it in my (own) writing."

"In my mind it looks like a book page."

"Numbers appear as on a blackboard."

"They're right smack in front of me."

"I can clearly see figures in my mind."

"Anything that follows a pattern is easy to do."

"I always had good teachers."

"I read the sports page all the time and figure team and player averages in my head."

"It's funny that I can get them, because I never really liked arithmetic."

By Low Achievers:

"Numbers won't stay."

"Can't remember."

"Hard to see it in my mind."

"Can't remember figures."

"Can't remember figures by hearing them only once."

"It just doesn't stick in my mind."

"I've never liked arithmetic since the fifth grade."

"I didn't try 'cause I didn't care."

"I fall behind trying to find an easy way."

"I didn't have enough time."

"We always moved from one school district to another."

"Math isn't my line. I like music."

The comments above seem to point to other possible characteristics of high achievers compared with low ones:

1) They seem to visualize better.

2) They seem to remember longer.

3) Through associations or something inherent

in their brain structure, stimuli became more intense or vivid for them.

4) They are more highly motivated, tend to enjoy their work.

D. Thought Processes and Methods of Solving Subtraction Examples in Relation to the Factors of School Grade, Sex, Intelligence, Ability in General Arithmetic, and Scores of Pupils on a Memory Span Test.

1. In relation to school grade: Older pupils tended to use a greater variety of methods of work, to choose a method more appropriate to the type of example for which solution was to be found.

2. In relation to sex: Though among the 143 high and low achievers (72 high and 71 low), 73 were girls and 70 boys, it is noteworthy that among the high achievers 61.1 per cent were boys and 38.9 per cent girls. Thirteen per cent of the boys varied their general methods (decomposition, carrying, etc.) to suit the type of subtraction example or type of presentation (oral, flash card, etc.); only seven per cent of the girls did so. Sixty-six per cent of the boys but only forty per cent of the girls varied their approaches (such as left-to-right, rounding, etc.)

3. In relation to general intelligence: The average intelligence quotient of the high achievers was 115.4 and that of the low group 99.9. The average of all of the pupils in the study was 105. The average of the low achievers would have been lower but for the fact that no one was selected for interviewing whose IQ fell below 90. This decision was made because it seemed desirable to study achievers whose poor performance to a large degree might be attributable to their ineffective methods of work rather than to their low intelligence.

Further study of the pupils revealed that the average IQ of those who used methods especially suited to mental arithmetic--left-to-right, rounding, place-grouping, and number-relations approaches--was 117. The mean of those who failed to use such methods was 107. Intelligence thus seems to have been an important factor in thought processes.

4. In relation to general arithmetic achievement: In Grades 6 through 9 in which this factor was studied, the average arithmetic grade score, based upon the arithmetic section of the Stanford Achievement test, of the high achievers was 9.4. That of the low group was 6.6. It is clear, therefore, that general arithmetic ability is a factor in mental arithmetic. Further evidence is the fact that the mean arithmetic score of those who used two or more methods for thinking subtraction depending upon the need and ease of solution was 9.2, that of

those who clung consistently to one method, 7.1.

5. In relation to memory span: Based upon a memory for digits test, having a range of possible scores from 0 to six levels, the mean level of the high achievers was 5.3, that of the low performers 4.1. The mean for those pupils who used two or more methods of solving the subtraction examples was 5.1; the mean of those who used only one basic method was 4.2. Thus, even though the correlation between memory span and mental arithmetic was found to be only .35 (mentioned earlier), the highest achievers in mental arithmetic appeared to surpass also in memory span.

Conclusions

Some of the more important conclusions that may be drawn from this study are the following:

A. General Conclusions.

1. Mental arithmetic, which in this study has been referred to as the mental or thought processes pupils engage in when attempting to solve arithmetic examples or problems without the use of paper and pencils, is facilitated by means of flashcard presentation as over against oral presentation. Also, pupils in Grades 6 through 12 can more readily solve subtraction examples when they have an opportunity to register their thought processes or work on paper with a pencil. In this study, pupils scored 50.4 per cent on a subtraction test presented with flash cards, and 88.9 per cent when paper and pencils were used.
2. After the eighth grade there was only a small gain in ability to solve subtraction examples without paper and pencil.
3. Boys excelled girls in mental arithmetic though girls were at least as good as boys when paper and pencil use was permitted.
4. A correlation of .50 was found between ability in mental arithmetic and intelligence, and a correlation of .65 between the former and general arithmetic achievement as measured by the Stanford Achievement test.
5. A low correlation of .35 was found between memory span for digits and mental arithmetic. This finding was contrary to expectations.

B. Conclusions Based Upon a Study of High and Low Achievers in Mental Arithmetic.

1. More than 90 per cent of the combined two groups of pupils used the "take-away" method of thinking subtraction and also the decomposition or "borrowing" procedure.
2. In addition to the "take-away" and additive methods and the "borrowing" and "carrying" procedures, use was also made of six additional thought processes that have been

labelled as Counting, Right-to-left Approach, Left-to-right Approach, Rounding, Place-Grouping, and Number-relations Approach. The last four approaches, particularly, seem to be associated with mental arithmetic.

3. Low achievers tended to use counting and the prevalent right-to-left approach which is stressed in most schools. High achievers tended to employ the left-to-right approach, rounding, place-grouping, and number-relations approach.
4. High achievers tended to use a greater variety of thought processes in mentally carrying on subtraction, ostensibly to fit the nature or difficulty of the subtraction problem. Poor achievers were inclined to stay with one basic method.
5. Many more high than low achievers began mentally solving the subtraction example when the first figure of the subtrahend was heard during the oral presentation. This of course implies that use was being made of the left-to-right approach.
6. In the solution of examples, rounding numbers which characterized particularly the work of high achievers, was carried on either by rounding downward from the minuend to the subtrahend or the reverse. Twice as many rounded upward. However, no advantage was observed in the one or the other of these two directions of work.

Implications

The implications that probably might be drawn from these findings are:

1) Since high achievers in mental arithmetic use methods of work usually not employed by low achievers, there probably are advantages in these methods. To the extent that poor achievers can understand and employ these methods, it would seem desirable that they be encouraged to use these processes.

2) Most schools do not seem to teach the thought processes especially suited to mental arithmetic, such as the left-to-right approach. Perhaps we should not leave to the accident of pupil resourcefulness or initiative the acquisition of these methods of work. Instead, we probably should provide for them in the arithmetic program.

3) Though the pupil's maturity, sex, intelligence, and general arithmetic achievement are particularly associated with facility with mental arithmetic in subtraction, it is possible that good teaching may at least in part compensate for the operation of these factors.

REFERENCES

1. Boulware, C. Elwood, The Emerging Concept of Mental Arithmetic, Unpublished Ed. D. dis-

- sertation, Teachers College, Columbia University, 1948.
2. Flournoy, Mary Frances, A Study of the Effectiveness of an Oral Arithmetic Program Prepared for Use at the Intermediate Grade Level, Unpublished Ph. D. dissertation, State University of Iowa, 1953.
 3. Hall, Jack V., The Ability of Elementary Teachers in Training and Sixth Grade Children to Solve Certain Arithmetic Problems Mentally and in Any Manner, Unpublished Ed. D. dissertation, Colorado State College, 1951.
 4. Petty, Olan K., A Study of Non-pencil and Paper Methods of Solving Arithmetic Word Problems Presented Visually, Unpublished Ph. D. dissertation, State University of Iowa, 1952.

SEX DIFFERENCES AMONG HIGH SCHOOL SENIORS

PAIRLEE J. STINSON, St. Louis University
MILDRED M. MORRISON, Maplewood, Missouri

MUCH CRITICISM is currently being made of the content of our high school curricula. Many new and sometimes unrealistic demands have been made of our teachers and administrators. These and other "sputnik-oriented" problems have made the job of the high school counselor very difficult, but challenging. The counselor now, more than ever before, must be effective in guiding and directing young people to equip themselves to cope and deal intelligently with the problems of the so-called scientific age.

The fundamental objective of counselors is to help students develop and grow intellectually, socially and emotionally to their maximum capacity. If this objective is to be realized, individual differences among students must be identified and analyzed early in the high school career. Each student should be directed in the light of his particular pattern of interests, abilities, and personality characteristics. The counselor must be able to predict with a high degree of accuracy whether a student will be able to succeed, i.e., perform effectively and be happy in a particular academic program, profession or vocation.

Accurate predictions demand more concise measurements and a more comprehensive understanding of the significance of these measurements. Much has been done to improve existing methods of making predictions about high school students, but current techniques and procedures are still far from satisfactory.

The measurement problems in the high schools are many and varied. One of the most difficult problems facing the high school counselor is concerned with selecting tests that will provide him with the maximum amount of valid and reliable information about his students at the minimum amount of cost, time for administration, and complexity of interpretation. Also the selected tests must be amenable to effective utilization by teachers and other school personnel who have had little or no special training in the field of measurement. This study is concerned with one approach to the test-selection problem.

Description of Tests

The problem of this study is to investigate the relationship of the subtests of the Differential Aptitude Tests (hereafter referred to as the DAT) and the Cooperative English, C2 (hereafter referred to as the Coop. Reading) to grade point average and to an abbreviated form of the Wechsler Adult Intelligence Scale (hereafter referred to as the WAIS). The DAT consists of the following eight subtests: 1) Verbal Reasoning, 2) Numerical Ability, 3) Abstract Reasoning, 4) Space Relations, 5) Mechanical Reasoning, 6) Clerical Speed and Accuracy, 7) Spelling, and 8) Sentences. The latter two comprise the Language Usage section of the battery and are considered to be more nearly achievement tests than any of the others. The first three tests, according to the authors (1), measure those functions which are associated with general intelligence. Space Relations and Mechanical Reasoning measure the ability to visualize and manipulate three dimensions in a two-dimensional surface and to recognize physical forces and principles. Clerical Speed and Accuracy measures the ability accurately and rapidly to perceive and respond to detail.

The Coop. Reading, C2 measures three aspects of reading: vocabulary knowledge, speed of reading, and level of comprehension. In addition to these specific scores, it yields a total weighted score. The C2 form of the test was selected because it is designed for use at the college level and with junior and senior high school students, particularly those who are somewhat advanced.

Hypotheses to be Tested

The investigators had made certain hypotheses about the relationships between the variables, and purposed, on the basis of their findings, to recommend a very brief battery of tests which would accomplish the same purpose as the rather lengthy one. They were further interested in studying sex differences that might exist among seniors in a rather typical high school. The principal hypotheses

to be evaluated in this investigation are as follows:

1. Verbal Reasoning, Numerical Ability, and Abstract Reasoning correlate more highly than other DAT subtests with general intelligence as evidenced by abbreviated WAIS scores.

2. Spelling, Sentences, and Verbal Reasoning correlate highly with grade point averages.

3. Reading comprehension correlates higher than any of the DAT subtests or total reading with WAIS IQ and grade point average.

4. Significant sex differences exist between scores on Verbal Reasoning (higher for girls), Space Relations (higher for boys), Mechanical Comprehension (higher for boys), and both Language Usage tests (higher for girls). In other words, it was hypothesized that the girls would score significantly higher on the verbal tests and the boys would perform better on some of the non-verbal or manipulative tests.

Method

The subjects of the study consist of 36 boys and 33 girls chosen at random from the senior class at the Maplewood-Richmond Heights Senior High School in Maplewood, Missouri. The ages of the subjects range from 16 years 8 months to 19 years for boys with a mean of 17 years 5 months; and from 16 years 5 months to 18 years 10 months for the girls with a mean age of 17 years 3 months. The DAT (Form A) and the Coop. Reading (Form Z) were administered to the subjects as a group in the fall of 1957. Each subject was then given an abbreviated form of the WAIS during a six week period following the group testing.

Research with short forms of the WAIS (2) indicate that valid measures of intelligence may be obtained by using well-selected subtests. The short form used in this study consisted of the following three verbal and three performance subtests: Information, Similarities, Vocabulary, Digit Symbol, Picture Completion, and Block Design. The total IQ was obtained by the standard procedure.

The grade point average for each student was computed by assigning numerical weights to the latter grades; thus, A+ = 12, A = 11, A- = 10, --D- = 1, and F = 0. The grade point average was based upon the accumulation of grades in all courses taken during the first three years of high school.

Correlational analysis was used to establish the relationships between the variables. The DAT and Coop. Reading were considered predictor variables and the WAIS and grade point average were the criterion variables. Pearson product-moment coefficients of correlation were calculated to estimate the relationship of each subtest of the DAT and Coop. Reading to WAIS IQ and grade point average. The over-all correlations among the variables were obtained by using the Fisher z-transformation technique (3). The t-test was used to determine the significance of the difference between the means of the two sexes on each variable.

Presentation and Discussion of Results

The correlations of the two criterion variables with each of the predictor variables are presented in Table I. The following general conclusions may be drawn from the results and evidence reported in this Table:

1. The relationship between the test variables and IQ is consistently higher for both sexes than the relationship between the test variables and grade point average. This suggests that factors other than those being measured by the test variables are operating in determining grades.

2. The WAIS and Coop. Reading correlate more highly with grade point average than any other single variable. This does not preclude that a selected weighted battery might not produce a higher relationship.

3. The relationships are consistently higher for boys except in three instances. Three of the DAT tests (Verbal Reasoning, Abstract Reasoning, and Space Relations) correlate higher with both WAIS and grade point average for girls than for boys. In general it might be said that girls are not as predictable as boys in this particular sample of the senior high school population. One might offer several explanations for this, but it is beyond the scope of this paper to do so. Undoubtedly, this is indicative of a difference in the manner of grading between boys and girls. Also, this does not indicate that grades in specific courses might not be more accurately predicted for girls than for boys using these same predictor variables.

4. The total reading score correlates .86 (boys), .84 (girls), and .85 (both sexes) with the WAIS. These are the highest correlations obtained in the entire study. As was previously noted, the reading comprehension score, with the exception of the WAIS, correlates higher with grade point average than any of the other tests. These two findings suggest that the Coop. Reading test is the best single test in the battery for a counselor to use in estimating a student's general level of intelligence and in predicting general academic success. The total score, which reflects vocabulary knowledge, speed of reading, and level of comprehension, seems to be very highly related to general intelligence. Further, the relationship is the same for both sexes. The comprehension score appears to be better than the total score in predicting grade point average, and this relationship is higher for boys than for girls.

The means and standard deviations of each test variable and grade point average for boys and girls separately are shown in Table II.

Percentile ranks are indicated for the mean scores of each DAT and Coop. Reading test. The DAT percentiles are based on norms for twelfth grade boys and for twelfth grade girls respectively. The Coop. Reading percentiles are based on entering college freshmen norms.

It is readily noted that the mean IQ for boys is 110.94 and for the girls is 107.00. Although this

TABLE I

CORRELATIONS OF WAIS AND GRADE POINT AVERAGE WITH DAT AND
COOP. READING FOR BOYS AND GIRLS SEPARATELY

(N _{boys} = 36)				(N _{girls} = 33)		
Test Variables	WAIS			Grade Point Average		
	Boys	Girls	Ave. *	Boys	Girls	Ave. *
DAT						
Verbal	.65	.70	.68	.39	.49	.45
Numerical	.64	.50	.58	.62	.46	.55
Abstract	.52	.69	.61	.30	.37	.34
Space	.63	.68	.66	.42	.53	.48
Mechanical	.70	.44	.61	.49	.36	.50
Clerical	.37	.31	.35	.37	.30	.34
Spelling	.51	.32	.42	.53	.32	.43
Sentences	.73	.54	.64	.68	.45	.58
Coop. Reading						
Vocabulary	.71	.70	.71	.40	.41	.41
Speed	.85	.82	.84	.67	.51	.60
Comprehension	.85	.75	.80	.73	.57	.66
Total	.86	.84	.85	.64	.55	.60
				.74	.63	.69
WAIS						

*Computed using Fisher's z-transformation technique.

TABLE II

MEANS, STANDARD DEVIATIONS, PERCENTILES^f, AND T-TEST DATA FOR THE VARIABLES(N_{boys} = 36)(N_{girls} = 33)

Test	Boys			Girls			Difference Between Means	t-Values
	M	SD	%ile	M	SD	%ile		
DAT								
Verbal	31.20	13.15	60	28.5	10.10	53	2.70	1
Numerical	26.20	14.55	60	20.1	9.43	57	6.10	2.08*
Abstract	37.70	13.53	74	34.9	7.97	65	2.80	1.06
Space	64.13	25.73	60	53.3	21.38	65	10.83	1.90
Mechanical	48.08	9.55	51	29.1	12.31	55	18.98	7.12***
Clerical	52.08	11.03	31	61.1	13.07	42	9.02	3.08**
Spelling	53.11	23.14	50	62.6	9.39	38	9.49	2.26*
Sentences	35.14	18.09	42	40.7	13.71	30	5.56	1.45
Coop. Reading								
Vocabulary	51.16	8.60	27	47.1	7.27	15	4.06	2.13
Speed	55.25	12.94	35	52.8	10.04	28	2.45	1
Comprehension	57.55	9.45	49	54.1	6.16	36	3.45	1.81
Total	54.72	10.41	38	51.2	7.73	24	3.51	1.60
WAIS	110.94	11.61		107.00	8.75		3.94	1.60
G. P. A.	6.45	2.45		7.48	1.82		1.03	1.99*

^f Norms for DAT percentiles are 12th grade boys and girls respectively.

^f Norms for Coop. Reading percentiles are entering college freshmen.

* Significant at the .05 level of confidence.

** Significant at the .01 level of confidence.

*** Significant at the .001 level of confidence.

difference is not statistically significant it may, to some extent, account for the higher means for boys on most of the other test variables. Only on Clerical, Spelling, and Sentences do the girls have a higher mean (based on raw scores) than do the boys. When considering percentile ranks of the mean scores, the girls exceed the boys only on Space, Mechanical, and Clerical.

It appears that these boys and girls as a group compare quite favorably with high school seniors throughout the country on all DAT subjects except Clerical, Spelling, and Sentences. The latter two tests, as previously mentioned, are considered to be achievement tests. The relatively low percentile scores on the clerical test may reflect general background training where emphasis has been placed upon accuracy rather than speed. The relatively low percentiles on Coop. Reading do not adequately reflect reading competency since they are based on entering college freshmen norms. On the basis of the norm data provided in the manual (4), one might infer that the reading scores for the boys are average, but the girls probably do not read as well as the average high school senior.

It was hypothesized that significant differences would exist between the two sexes on various tests. These hypotheses were evaluated by using the t-test technique. Table II presents the t-values and their levels of significance. It is immediately apparent that most of the hypotheses regarding sex differences had to be rejected. Only on Mechanical and Spelling were there significant differences in the direction that was anticipated by the investigators. Boys demonstrate much more mechanical knowledge than girls. Girls, on the other hand, spell significantly better than the boys, although they do not compare as well as boys on national norms.

It was further hypothesized that the girls would have more capacity in reading skills than the boys. This hypothesis also had to be rejected. As a matter of fact, the boys' mean scores are higher on all four aspects of reading, although the difference is significant only on vocabulary knowledge. Of course, it must be constantly kept in mind that differences in intelligence (although not significant) might, in part, account for some of the variation in the other test scores.

Perhaps one of the most significant results of the study is revealed in the comparison of grade point averages for the two sexes. The girls obtained significantly higher grades than the boys. The mean grade point average for the girls in this study is between a B and a B-, whereas the boys' mean is between a C+ and a B-. In view of the fact that results indicate that the boys in the study are somewhat brighter and they demonstrate slightly better capacity on most of the aptitude tests, it would seem that they should get better grades than the girls. This again suggests that many "non-intellectual" factors operate in determining success in academic situations. Undoubtedly some of these "non-intellectual" factors might

include such things as general attitude toward school and the learning environment, study habits, interests, drive and motivation, home and family relationships, general health, and interpersonal relationships.

Summary and Conclusions

The purpose of this study was to investigate the relationship of general intelligence (WAIS) and grade point average to the subtests of the DAT and Coop. Reading, and to determine sex differences that might exist between mean test scores for a sample of high school seniors. Correlational analysis and t-test results evoked the following major conclusions:

1. The WAIS correlated higher between the test variables than did grade point average.
2. The correlations for boys were higher than the corresponding correlations for girls in most instances.
3. The highest relationships were found to be between each component of the Coop. Reading and the WAIS. The Coop. Reading correlated higher with both WAIS and grade point average than did the DAT.
4. The sample of subjects compared quite favorably with other high school seniors throughout the country except in Clerical, Spelling, Sentences, and, perhaps, reading skills.
5. Significant sex differences were found between mean scores on Numerical Reasoning (higher for boys), Mechanical Reasoning (higher for boys), Clerical Speed and Accuracy (higher for girls), Spelling (higher for girls), Coop. Reading, vocabulary (higher for boys), and grade point average (higher for girls).

It was suggested that, if a counselor wanted to use a minimum of tests, the Coop. Reading would supply him with the maximum amount of information about his students. The total reading score correlates about .85 (average for both sexes) with general intelligence. The comprehension score seems to be the highest predictor of grade point averages (.66 for both sexes). It, as well as the other tests, seems to predict academic success or behavior for boys much better than for girls. This lead the investigators to wonder, "Are Women Unpredictable?"

REFERENCES

1. Bennett, G., Seashore, H., and Wesman, A., A Manual for the Differential Aptitude Tests. New York: Psychological Corp., 1952.
2. Doppelt, J. E., "Estimating the Full Scale Score on the Wechsler Adult Intelligence Scale from Scores on Four Subtests." Journal of Consulting Psychology, 1956, 20, 63-66.
3. Garrett, H. E., Statistics in Psychology and Education. New York: Longmans, Green and Co., 1954.

4. The Cooperative Tests, Manual of Directions. New Jersey: Cooperative Test Division, Educational Testing Service, 1951.
5. Wechsler, D. , Manual for the Wechsler Adult Intelligence Scale. New York: Psychological Corp. , 1955.

PREDICTION OF SUCCESS IN GRADUATE SCHOOL AT RUTGERS UNIVERSITY*

MILTON M. SCHWARTZ and F. EUGENE CLARK
Rutgers University

THE PRESENT STUDY is part of a broad and continuing program of research on graduate student selection at Rutgers University. Although this paper concerns the Department of Applied and Mathematical Statistics, several other departments are also cooperating, namely the departments of English and Chemistry and the schools of Engineering and Agriculture. These results will be presented in forthcoming articles upon their completion.

The Department of Applied and Mechanical Statistics offers courses leading to the Master of Science degree for part-time evening students employed in local industry. Most of these students have returned to academic study after a lapse of several years. For this reason, it was felt undergraduate grades did not have the predictive significance they would have for recent graduates and another predictor was desired, preferably in the form of a convenient test. A total of 150 students enrolled in courses given by this department were tested and complete data for this study were finally obtained for 56 of these students. This constituted the experimental group, all of whom had been accepted for graduate study.

Three predictors of success were used, namely, undergraduate grade averages and two tests of The Psychological Corporation, restricted to use in graduate student selection, the Miller Analogies Test (1) and the newer Doppelt Mathematical Reasoning Test (2).² Three criteria of success were employed. They consisted of graduate grades, an academic competence rating, and a job-effectiveness rating. Further description of the predictors and criteria is given in the later sections of the article. In summary, three types of data were gathered: 1) Doppelt and Miller tests administered during the regular classroom period, 2) undergraduate and graduate grade averages and 3) rankings by instructors of academic competence and job effectiveness.

Predictors

The Miller and Doppelt tests consist of 100 and of 50 multiple-choice items respectively and employ a fifty-minute time limit. Both tests deal with the student's ability to discern relationships in items presented in the form of analogies. The Miller test deals with a wide variety of subject-matters and depends largely on the student's supply of factual information. The Doppelt test, not yet as widely employed, is designed to test the student's understanding of elementary mathematical concepts. Ability to utilize these commonly known concepts tests the student's mathematical reasoning ability. Thus it was felt that this instrument might prove to be an effective predictor for students of statistics.

The cumulative average for all undergraduate courses for each individual was available for use in this study. This score is based on the average of course grades weighted by number of credit-hours where the course grade is expressed by the number 1 for a grade of "A", 2 for "B", 3 for "C", 4 for "D" and 5 for a grade of "F". In the study the correlations between pairs of these predictors were computed. Correlation between Miller and Doppelt was .40 while that between undergraduate grades and either of the tests was near zero. In explanation of the rather startling lack of correlation with the undergraduate averages, it was felt that the diversity of standards among different institutions was probably an important factor. In addition, it is probable that a number of bright students did not do well when they were undergraduates for lack of motivation.

Criteria for Success

Graduate averages were computed in the same manner as undergraduate averages.

No attempt was made to define "academic com-

*Footnotes will be found at the end of the article.

TABLE I
DISTRIBUTION OF RANKINGS FOR EXPERIMENTAL GROUP

Number of Rankings	Number of Students Receiving Indicated Number of Rankings
	10
2	9
3	14
4	11
5	9
6	<u>3</u>
Total	56

TABLE II
CORRELATIONS BETWEEN PREDICTORS AND CRITERIA*

	Graduate Averages	Academic Rating	Job Rating
Miller	.26	.25	.25
Doppelt	.32	.44	.31
Undergraduate Average	.24	.18	.22
Miller and Doppelt	.35	.45	.32
Doppelt and Undergraduate Average	.41	.48	.38
Miller, Doppelt and Undergraduate Average	.43	.50	.41

* All correlations are product-moment.

petence" or "job effectiveness" beyond the information suggested by their labels and instructions for ranking supplied by the instructions. Judgment of academic competence was based essentially on conventional standards of classroom performance whereas the principle basis for judging a student's effectiveness on the job was his ability to transfer principles and knowledge asquired in the classroom to problems faced in the industrial setting. Students were required, as part of their courses, to submit term papers indicating such application.

The instructions for ranking were as follows:

"The attached list of students includes the names of those currently enrolled in courses in Applied and Mathematical Statistics. Will you please help us by ranking your students as follows:

1. Underline the names of all students you have had in class. If you don't recall the student well enough to have confidence in your judgment, also place a question mark beside the name.

2. Rank the students by quartiles in each of the two categories A and B on the attached sheet by writing one-fourth of the names underlined in each box on the accompanying sheet. (It is suggested you first pick the students you wish to rank at the extremes.)

- A. Your estimate of the student's academic competence.

- B. Your estimate of the student's effectiveness as a quality control engineer or administrator."

Thus a quartile ranking by each of the eight instructors in the department was prepared for each of the two criterion measures. A score of 1 was assigned to students ranked in the upper quartile, 2 for the second quartile, 3 for the third quartile and 4 for the lowest quartile. The number of rankings per student varied, of course, and the average of the ranks to one decimal point was used. In the treatment of data, individuals with less than four ratings were differentiated from those with four or more ratings. However, in the analysis of these data the dichotomy was abandoned since no reliable difference between the two groups was found, and these data were therefore combined.

Reasonable precautions were taken regarding experimental controls. Each instructor working independently ranked his group without reference to scores on any of the remaining variables. Of course each rater had information about the grades

he gave the student and perhaps grades they received in other courses. Correlation between graduate grades and academic rating was .79; between graduate grades and job rating was .61.

Analysis of Data

Machine computation was available for computation and a multiple correlation of predictors with the criteria. The results are shown in Table II.

The Miller test is seen to be ineffectual as a predictor for this particular group of students. The Doppelt test was the best single predictor as measured by any of the three criteria and particularly as measured against the academic rating.

The combination of all predictors added very little to the total correlation. None of the correlations is high enough to give strong confidence in any of the predictors. It is suspected that one important reason for this is the previously mentioned gap in the academic career of most of the group. Their success appears to be governed more by motivation than by the factors tested. It was felt, however, that the Doppelt test could serve as a useful screening device. Examination of scatter-diagrams suggested that a score of 30 was a reasonable cut-off point.

FOOTNOTES

- *1. The writers wish to express their appreciation to Deans Marion A. Johnson and John L. Swink and Professor Ellis Ott, all of Rutgers, for helping to make this research possible; also to the Psychological Corporation for making available without charge the Doppelt Mathematical Reasoning Test.
2. While time schedules could not be arranged to include the widely used Graduate Record Examination, its eventual inclusion in this battery as well as batteries for the other groups is planned.

REFERENCES

1. Miller Analogies Test Manual, The Psychological Corporation, 1952. New York.
2. Doppelt Mathematical Reasoning Test Manual, Psychological Corporation, 1958, New York.

ACCEPTANCE AND REJECTION AS RELATED TO LENGTH OF SCHOOL ATTENDANCE

JACK FRYMIER*
Alabama Polytechnic Institute
Auburn, Alabama

This paper represents a study of student acceptance and rejection within a particular school situation as related to the length of time the students attended school together.

The author, who was a core teacher in a laboratory school, felt that something was amiss in his particular section of core. His class appeared to be split into two groups. As he studied the situation over an extended period of time there appeared to be some relationship between these groups and the length of time the students had attended school together. There seemed to be, on the one hand, a group which was composed primarily of those students who had attended that school together since kindergarten or early elementary school. On the other hand there seemed to be another group, less well defined, which was composed primarily of students who had entered this particular school at the ninth grade level, and who had not necessarily attended elementary school with those other students who also entered at that time.

The purpose of this study was to determine if there was any difference between the extent to which members of these two groups were (a) accepted or rejected, (b) pointed out as being members of power factions, and (c) pointed out as isolates.

The Procedure

This school had a maintained enrollment of approximately 500 students from kindergarten through twelfth. In grades one through eight there was one section of 30 students each, but from nine through twelve there were two sections that size for each grade level. Because the number of students was doubled at the ninth grade level, each upcoming eighth grade was split into two groups of 15 students each, and an additional 15 students from other local or outlying schools were added to these groups to make two sections of 30 each.

For the purpose of this study those students who

had attended the elementary grades at this particular school were considered "oldtimers." Those who entered at the ninth grade or later were considered "newcomers."

In order to collect data for the study the author devised a simple sociometric-type questionnaire. This was administered to five of the eight core sections in grades nine through twelve. There were 141 students involved, 63 "oldtimers" and 78 newcomers." The questionnaire was as follows:

1. List here the two persons you would prefer to sit between in core.
2. List here any persons you would prefer not to sit with in core.
3. List here the two persons you would select to be on a committee studying some unit in core with you.
4. List here any persons you would prefer not to work with on a core committee.
5. Do you feel that there are any "power factions" in your core class? (Small groups whose primary aim is to get their own way.) If so, would you list them below.
6. List here any persons you feel are isolates (neglected, ignored, not a part of the group).

Questions one and three were the only ones the students were required to answer. It was explained that numbers two, four, five, and six did not necessarily have to be completed unless the students felt that way about some particular person.

If a student's name was listed in response to question one or three, this was assumed to indicate some degree of acceptance. If a student's name was listed in response to question two or four, this was assumed to indicate rejection to some degree. Further, it was assumed that question five indicated recognition of extreme acceptance within the situation and question six recognition of extreme rejection.

* The author is indebted to Dr. Bernard Corman, Michigan State University, and Dr. Arthur Combs, University of Florida, for their assistance throughout the various stages of this problem.

For purposes of this study, then, it was assumed that questions one through four determined peer acceptance or rejection, whereas questions five and six served as an indication of student recognition of these factors within their core groups.

The results were tabulated as follows: each time a student's name was listed after either question one or three he received one plus score (plus interpreted as acceptance), whereas each time a student's name was listed in response to either question two or four he received one minus score (minus interpreted as rejection). The results of these four questions were added together for each student and the mean score and standard deviation for the "oldtimer" group and "newcomer" group computed. During the tabulation it was also noted whether or not this acceptance or rejection was by an "oldtimer" or by a "newcomer."

The total number of times a student's name was listed in response to question five was also recorded and the mean scores and standard deviation for the "oldtimer" and "newcomer" groups computed. The same was done for question six regarding isolation. By the nature of these computations the scores for questions five and six were all positive numbers, whereas the scores for questions one through four were both positive and negative.

The Results

The results of these computations are recorded in Tables I through III.

TABLE I

ACCEPTANCE AND REJECTION OF OLDTIMERS AND NEWCOMERS

Group	Mean Score	Standard Deviation
Oldtimers	+1.87	4.61
Newcomers	-.03	6.10

Note: Plus score indicates acceptance, minus score indicates rejection.

Table I shows the results of responses to questions one through four. "Oldtimers" were accepted (mean +1.87), whereas "newcomers" tended to be rejected slightly (mean -.03). Further analysis of the data show that "oldtimers" accepted other individuals within the "oldtimer" group (mean +1.19) but rejected "newcomers" (mean -.76). On the other hand, "newcomers" accepted "oldtimers" (mean +.61) and other "newcomers" (mean +.28). This seems to indicate that those persons who had

attended school together for an extended period of time were accepted to a greater degree than were the newer members of the group. Further, rejection came primarily from the "oldtimer" group rather than from both groups indiscriminately.

TABLE II

EXTENT TO WHICH STUDENTS WERE POINTED OUT AS MEMBERS OF A POWER FACTION

Group	Mean Score	Standard Deviation
Oldtimers	3.08	3.57
Newcomers	.94	1.50

Table II represents the data regarding question number five, power factions. Members of the "oldtimer" group were labeled as being members of power factions (mean 3.08) more frequently than were "newcomers" (mean .94). There apparently was something fundamentally different in the way these students functioned within their core sections to evoke such a recognized difference.

TABLE III

EXTENT TO WHICH STUDENTS WERE POINTED OUT AS ISOLATES

Group	Mean Score	Standard Deviation
Oldtimers	.73	2.26
Newcomers	2.65	4.01

Table III indicates that "newcomers" were pointed out as isolates (mean 2.65) much more frequently than were the members of the "oldtimer" group (mean .73).

A further examination of the raw scores indicates that of the eight most highly accepted students, four were "oldtimers" and four "newcomers." However, of the eight most severely rejected subjects, one was an "oldtimer" and seven were "newcomers."

Of the eight students most frequently mentioned as being members of a power faction seven were from the "oldtimer" group and one was a "newcomer."

Finally, of the eight persons listed most often as isolates, one was an "oldtimer" and seven were "newcomers."

Discussion of Results

There seemed to be considerable difference in the way students were accepted or rejected within a group, and there also was some evidence that this may have been related to the length of time they had been a member of the group. Many individuals coming into an established group apparently experienced considerable difficulty getting accepted. The data seem to indicate that incoming members may have been rejected simply because they had not been a part of the established group for any period of time.

It appears that if considerable number of indi-

viduals join an established group simultaneously, this action may cause certain forces to generate within the situation which results in the original members merging into a strong, cohesive unit, accepting their own members, but rejecting new incoming members without a great deal of consideration. This acceptance and rejection seems to exist and also seems to be perceived by the members of both groups quite accurately.

It would be interesting to determine whether or not acceptance or rejection is related in any way to academic achievement, student mortality, participation in extracurricular activities, or any other such activities.

SOME ATTITUDINAL DIFFERENCES AMONG EDUCATIONAL SPECIALISTS, ADMIN- ISTRATORS AND TEACHERS

BARTHOLOMEW D. WALL
Wayne State University, Detroit

Statement of the Problem

Educational literature contains numerous references to a purportedly "unsatisfactory" state of interpersonal relationships existing among educational specialists, administrators and teachers. (6). Some authorities trace this difficulty to attitudinal differences (5), although research to date has verified neither the existence of attitudinal differences nor the presence of "unsatisfactory" interpersonal relationships (4). The purpose of this investigation was to determine whether there exist significant attitudinal differences among educational specialists, administrators and teachers employed in the public elementary schools of Wayne County, Michigan. The study was conducted secondarily for the purpose of throwing some light upon the possible relationship of attitudinal differences to interpersonal relationships and upon the threat to the use of cooperative techniques among educational personnel.

The Instrument

The attitude scale or questionnaire used in this research consisted of a series of forty items, negatively worded, to which the respondent was asked to react on a basis of the following statements: 1) Agree, 2) Disagree, or 3) Undecided. Included in the scale were two pairs of items specifically designed to check on the respondent's consistency. An additional ten items, five borrowed from the F-Scale, were designed to measure the flexibility of the respondent.

The Attitudes to be Measured

Consensus as to what would constitute the most crucial kinds of attitudes to be measured was secured from a panel of ten educational experts. The attitudes to be measured fell under the following category titles:

I. Attitudes Toward the Profession, for the purpose of obtaining some measure of the individual's feeling of competency in and value for the profession.

II. Attitudes Toward Children, for the purpose of deriving some measure of the respondent's acceptance of personality growth and development theory, learning theory, pupil-teacher planning and discipline as inner controls over impulsivity.

III. Attitudes Toward One's Professional Peers, for the purpose of measuring the individual's perception of and acceptance of peer roles.

IV. Attitudes Toward Education as a Community Agency, for the purpose of measuring the respondent's understanding of and acceptance of the role of education as an institution having community responsibilities.

Method of Collecting Data

A mail questionnaire was employed to collect the data. The validity of the instrument was attested 1) by developing the scale on a basis of the consensus of a panel of educational experts, 2) by executing a personal interview follow-up study.

Pilot Study

Reliability data for the P-Scale was obtained by administering the P-Scale to two sections of graduate students, College of Education, Wayne State University, on a test-retest basis, thirty-five day interval. The reliability data were as follows:

1. 1st test, odd-even r.94
2. 1st test, split-half r.96
3. Retest, odd-even r.96
4. Retest, split-half r.97
5. Test-retest r.98
6. Rank-difference r.86

Operational Definition

Persons in the field of education function attitudinally along a continuum which can be termed democratic, liberal and progressive at one end and authoritarian, conservative and traditional at the other. It is possible to measure a person's place on such a continuum by using a valid and reliable atti-

TABLE I
RAW SCORES, MEAN SCORES, VARIANCES AND STANDARD
DEVIATIONS FOR THE THREE GROUPS

Group	Total Raw Scores	Means	Variances	Sigma
Specialists (N/30)	1433	48	31	5.568
Administrators (N/30)	2001	66	145	12.042
Teachers (N/30)	2283	76	222	14.900

TABLE II
MEAN SCORE DIFFERENCES FOR THE GROUPS' SCORES

Groups	T-Scores	Value of T at 58 °/ Freedom
Specialist-Administrator	7.29	2.663 1% level
Specialist-Teacher	10.11	2.663 1% level
Teacher-Administrator	2.81	2.663 • 1% level

tude scale in which each item falls upon the continuum and occupies relatively equal intervals. Toward this end, the panel of educational experts rated each item in terms of the operational definition, in terms of Parten's rules for the construction of scale items (7), in terms of Kilpatrick's rules for the elimination of unsuitable material (7), and in terms of Blankenship's list of danger words (1).

Attitudes, by their very definition, are unconscious and in order to derive some measurement of attitudes, one must measure a person's opinions, assuming for the purposes of the research that there is a very close relationship existing between opinions and attitudes.

Hypothesis

There may exist statistically significant attitudinal differences among educational specialists, administrators and teachers employed in the public elementary schools of Wayne County, Michigan.

Null Hypothesis

There exist no demonstrable attitudinal differences which are statistically significant among educational specialists, administrators and teachers employed in the public elementary schools of Wayne County, Michigan.

The Sample

Garrett states that minimum numbers would be 25 cases (3). Edwards indicates that 80 cases would constitute minimum numbers (2). The following numbers constituted the sample employed:

1. Educational Specialists.....	30
2. Administrators	30
3. Teachers	30
	90

Other groups of specialists were included in the survey but have not been included in the report. The sample was selected from lists of employees' names contained on rosters supplied the Wayne County Board of Education by local school districts in Wayne County, Michigan. Using double columns of the table of random numbers, the author selected the names in the quantities needed for the study. The names of personnel were listed 100 to the page and numbered. The specialists sample included 15 diagnosticians and 15 visiting teachers. The administrator sample included 15 principals and 15 superintendents. The teacher sample included 30 teachers, kindergarten through sixth grade. The school districts from which the names were drawn were first classified as urban, suburban or rural to insure representative coverage of the county.

Basic Data for the Three Groups

The educational specialists, administrators and teachers obtained raw scores, mean scores, variances and standard deviations as shown in Table I. The mean score differences for the groups' scores on the P-Scale were statistically significant as attested by the T-Scores shown in Table II.

Findings

The major finding is that statistically significant attitudinal differences exist among the mentioned specialists, administrators and teachers.

Minor findings were to the effect that the specialists in Wayne County appear to be more democratic, liberal and progressive in their attitudes as measured by the P-Scale than either administrators or teachers. The teachers appear to be more authoritarian, conservative and traditional than either the administrators or the specialists. Certain flexibility and consistency items were included in the scale and on a basis of their responses to those items, the specialists appeared to be more flexible and more consistent than either the administrators or the teachers.

The sub-scale or category score patterns were similar for all three groups indicating possibly that the differences are more of degree than of kind.

REFERENCES

1. Blankenship, Albert, "The Choice of Words in Poll Questions," *Sociology and Social Research*, XXV (1940), 12-18.
2. Edwards, Allen L., *Statistical Analysis*. (New York: Rinehart and Company, Inc., 1946, p. 166.
3. Garrett, Henry E., *Statistics in Psychology and Education*. (New York: Longmans, Green and Company, Inc., 1946), p. 225.
4. Koopman, G. Robert, "Teamwork and Social Change," *Educational Leadership*. (Washington, D. C.: Association for Supervision and Curriculum Development, 1957), XIV:6, p. 356.
5. Leonard, J. Paul, *Developing the Secondary School Curriculum*. (New York: Rinehart and Company, Inc., 1946), p. 77.
6. London, Jack, "The Development of a Grievance Procedure in the Public Schools," *Educational Administration and Supervision*. (Baltimore: Warwick and York, Inc., 1957), XLIII: 1, p. 4.
7. Parten, Mildred, *Surveys, Polls and Samples: Practical Procedures*. (New York: Harpers, 1950), p. 200.

INTELLECTUAL AND PERSONALITY CHARACTERISTICS OF UNIVERSITY OF UTAH STUDENTS

FRANK B. JEX, REED M. MERRILL
University of Utah

STAFF MEMBERS and graduate students in the Counseling Center and Department of Educational Psychology, University of Utah, have gathered a large array of data related to counseling and testing activities. These data have been distributed to the faculty, administration and professional colleagues in a series of Counseling Center reports*, and have formed the basis for a number of administrative and counseling procedures in the General Education program, the Admissions program and in other University of Utah programs. Following is a summary of some of the most interesting findings. Although the findings relate specifically to University of Utah students, it is felt that there may be considerable generality in the findings which are significant for other college student populations.

The typical entering freshman at the University of Utah has an intelligence quotient (IQ) of around one hundred and fifteen, which is about average for American college students in general, but somewhat lower than the average for state universities. The typical graduating senior has an IQ of around one hundred and twenty-five, although about one graduate in fifty has an IQ below one hundred. However, five-sixths of University of Utah graduates come from the top fifth of the general population, intellectually speaking, and very few students acquire higher professional degrees who do not have an IQ of at least one hundred and twenty.

In recent years it has been customary to break down general intelligence into its more specific factors. When the typical entering freshman is analyzed in this fashion, his strongest ability is reasoning, followed by spatial ability, verbal ability, numerical ability, and finally clerical ability (which is but little better than that for the adult population at large). Those who emerge as graduating seniors are strongest in the area of verbal ability, which suggests the crucial role of communications in the curriculum.

But even more important than general intellectual ability in determining who will succeed in this University are the student's high school record and his entrance test scores. Together they account for about half of what goes into college success as measured by grades. This raises the general question of the typical student's academic predictability.

Pehlke's (6) recent survey indicates that while the typical entering freshman thinks it is a good idea for us to administer entrance tests, so that he can make what is perhaps an "agonizing self-appraisal" of his scholastic potential, and thus be stimulated to do his best, he does not favor the use of these tests for admission purposes. The much more popular opinion is that all high school graduates should be allowed to enter college. This leads to what we may appropriately call our "wash-out" system. It means that while the typical freshman makes a passing average in his first quarter at the university, forty percent of his less typical friends fail to pass the first quarter.

Even more disconcerting is the fact that only about one in four of aspiring freshmen manages to graduate. The fact that about eighty-five percent of our graduates come from those freshmen who pass their first quarter in college attests to both the consistency of students' academic performance and the consistency of their professors' grading practices. As a matter of fact, the entrance test data, combined with the high school average, pretty well indicate at the outset whether or not a person will succeed in college. If he earned good grades in high school and achieved a high level of reading, writing and arithmetic skill, he will most likely continue this pattern of success into college. Based on these criteria of high school average and entrance test scores, the research indicates that no more than ten percent of the bottom fifth of an entering class can be expected to graduate, while more than forty percent of the top fifth will graduate.

* For those interested in a more detailed analysis, the reports can be obtained by writing the Director, Counseling Center, University of Utah, Salt Lake City 12, Utah.

What, then, is the high school background and entrance test profile of the entering freshman student? Typically he was a "B" student in high school, with his strongest skills in the natural sciences, and his weakest skills in English, particularly in the mechanics of expression. Actually, his grammatical usage, his punctuation, his spelling, and his sentence structure are all considerably poorer than that of the average American high school graduate. This takes on considerable significance when it is known that these English mechanics scores are more closely related to college success than scores on any of the other entrance tests. Add to this the fact that, of the entering freshmen who were no better than "C" in high school, not more than one in ten succeeds in graduating, and the seriousness of a situation which attracts to college so many youths with lower level verbal ability and poor scholastic records becomes apparent. What is it that attracts some fifty percent of our high school graduates to college? Undoubtedly, much of it is due to a long tradition which from its founding has viewed the University of Utah as an indispensable capstone to the public school system of the state. In Utah it is expected that you go on to college, even though you may have resisted high school education with considerable success.

But what are the typical student's reasons for coming to college? The local analyses reveal that vocational considerations are clearly dominant. College is seen as the open sesame to material well-being. There is much bickering among educators as to whether the colleges should be engaged primarily in teaching their students how to live or how to make a living. But our students know what they want. They have come to prepare themselves for better paying jobs. Clearly, they seek not so much the essence of education as its symbols: the grades, the degrees and the certificates. They move in increasing numbers to the professional colleges in which sixty percent of them are currently enrolled. They are after degrees that will pay off immediately upon graduation. Less compelling reasons for coming to college are to pursue their academic interests, to increase their opportunities for social service, and for personal improvement, in that order.

One of the most interesting sidelights to this very materialistic orientation is the fact that so many of the students are gainfully employed a substantial part of their time while carrying a full college load. Various local studies show that from fifty to seventy percent of college men are working on the average twenty-two hours a week, although Lloyd's (5) recent findings indicate that only one in ten of those who work over twenty-four hours a week will graduate. It is suspected that the somewhat jaundiced eye of the college professor sees a typical college student who is so busy working, and dating, and mating, that there is precious little time left for cerebrating.

A further breakdown of our data reveals some

highly significant sex differences. Women students achieve higher scores in English entrance tests, especially in English mechanics, but are poorer on the other entrance tests. Women students receive higher grades in high school by about one third of a grade point and higher grades in college by about a quarter of a grade point. In spite of the general impression to the contrary, women are considerably more predictable with respect to grade-getting than are men. On the other hand, men do better on the entrance tests in social studies, natural sciences and mathematics, and on tests of general scholastic aptitude. From which it might be inferred that, academically speaking, women do more with less (which many of their professors have long suspected).

In speculating as to the significance of these sex differences, it seems reasonable to assume that since fewer women than men come to college, additional selective factors operate to determine which women will attend. From other evidence which has been accumulated, it appears that our college women are scholastically more conforming and more highly motivated. By avoiding the technical scientific areas for which they are not as well prepared, and concentrating on the more verbal areas for which they are much better prepared, the women are able to "out-grade-point" the men by a considerable margin.

As would be expected, the studies related to interest and personality are much less definitive than those of intellectual and academic characteristics of students. Stott (11) compared a group of University of Utah male students with male students from Oregon and California on the F-Scale measure of the authoritarian personality. The University of Utah students were more authoritarian, more ethnocentric and more politically and economically conservative than the Oregon and California student samples. The F-Scale defines the authoritarian personality as a "general tendency to glorify, to be subservient to and remain uncritical toward authoritarian figures in the name of some moral authority". It is quite likely that the findings are related to the unusual cultural history of the state of Utah. Related to this is a population statistic that makes the University of Utah unique among state universities. Pennock (7) found that sixty-nine percent of the student body stated that their religious affiliation was with the Church of Jesus Christ of Latter Day Saints. Probably no other state university has such a large single religious group in its student body.

Gaines (2) studied the hypothetical "ideal" college graduate as described by university faculty and successful business men. Overall there was much agreement in these descriptions. However, there are several areas of significant difference; the faculty being more concerned than the business man in the "ideal" student's independence, objectivity, and humanistic motivation and ethics.

Pennock (7) studied the sexual attitudes and be-

havior of a sample of college men. His findings indicated that the sexual behavior of University of Utah males was very similar to the findings for such a group reported in the Kinsey studies. Gardner (3) studied college males using as a measure of personality the Minnesota Multiphasic Personality Inventory. His findings were very similar to studies utilizing this instrument on other college campuses. That is, college males are more active, less concerned about masculinity-femininity differences and less conforming than the general population as measured by the MMPI.

Cohen (1) investigated the student who had considerable academic promise but managed to accomplish a straight "E" record. This is a small but somewhat dubiously select group of students. There are many interest and personality factors related to the problems of this group but the outstanding characteristic of the "E" students of academic promise seemed to be that they lacked the motivation to succeed. They do have the motivation to fail. It was concluded for this group that failure was a purposeful, planned activity.

Sorenson (12) studied the relation of non-intellectual factors to success in college. His findings in regard to bright-passing and bright-failing men are of interest. Bright-passing men are characterized by having better family relations and better social adjustment during high school. They recognize their emotional states and admit having minor physical ailments. They have an aptitude and interest for English. They report a lack of cheating on high school examinations. They are more likely to say they came to college "to get an education." More of the passing group participated in high school extra-curricular activities and they have a greater conformity with respect to smoking, drinking, attitudes toward sex, marriage and church attendance. The opposite is reported for bright-failing men with the addition that more of them reported feeling shy when meeting people for the first time.

Several studies have indicated that the most seriously motivated students average the highest grades and the least seriously motivated, the lowest grades. Research here and elsewhere does not support the commonly held opinion that amount of study is a crucial factor in the achievement of grades. Apparently, what a student actually does in the library is still more important than whether he is simply there or not.

There is much evidence to defend the point of view that, in general, university students are above average members of our society, intellectually, personality-wise, and occupationally. Nonetheless,

the university community includes students, and no doubt faculty, who represent the entire gamut of personality and interest characteristics found in the general population.

REFERENCES

1. Cohen, Daniel J., Factors Related to the Academic Failure of Students Who Were Predicted to Succeed in College. Master's thesis, University of Utah, 1955.
2. Gaines, Grant R., A Q-Technique Study of Characteristics in the Hypothetical 'Ideal' College Graduate. Doctoral dissertation, University of Utah, 1956.
3. Gardner, David Bruce, Norms for College Students for the Minnesota Multiphasic Personality Inventory. Master's thesis, University of Utah, 1949.
4. Jex, Frank B., "University of Utah Studies in the Prediction of Academic Success." University of Utah Research Monographs in Education, Vol. 1, July, 1957.
5. Lloyd, Charles P., A Study of Scholastically Promising Students Who Withdrew from College Prior to Graduation. Master's Thesis, University of Utah, 1955.
6. Pehlke, Donald E., The Orientation of Freshmen in State Universities. Doctoral dissertation, University of Utah, 1952.
7. Pennock, John Albert, A Study of the Sexual Attitudes and Behavior of 200 Single College Men. Master's thesis, University of Utah, 1949.
8. Sharp, Heber C., A Comparative Analysis of the Ohio State Psychological Scores of the Yearly Groups of University of Utah Freshmen. Master's thesis, University of Utah, 1945.
9. Senior, Noel, An Analysis of the Effect of Four Years of College Training on General Aptitude Scores. Master's thesis, University of Utah, 1952.
10. G. A. T. B. Senior Project Staff, Department of Psychology, University of Utah, "G. A. T. B. Patterns for College Areas," Occupations, April, 1951. pp. 518-526.
11. Stott, Sterling S., A Comparison of the Authoritarian Personality F-Scale with the Allport-Vernon-Lindzey Study of Values. Master's thesis, University of Utah, 1954.
12. Sorensen, A. Garth, Non-Intellectual Factors Related to Success in College. Doctoral dissertation, University of Utah, 1950.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 53, NUMBER 4, DECEMBER 1959

CONTENTS

Fitting Asymmetric Student Grade Distributions—Robert H. Riffenburgh	123
Socioeconomic Status, Urbanism and Academic Perform- ance in College—Norman F. Washburne.....	130
The Relationship of Selected Measures to Spelling Achievement at the Fourth and Eighth Grade Levels— I. E. Aaron.....	138
Merit-Rating Salary Plans in Public School Systems of the United States, 1955-56—Irvin Albert Karam.....	144
An Investigation of the Effects of a Seventh and Eighth Grade Core Program—Bernard Schwartz.....	149
A Study of Student Disciplinary Practices in Two Geor- gia High Schools—Karl C. Garrison.....	153
Some Results of an Enrichment Program for Gifted Ninth Graders—J. A. R. Wilson	157

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

FITTING ASYMMETRIC STUDENT GRADE DISTRIBUTIONS

ROBERT H. RIFFENBURGH
University of Hawaii

IT IS THE PURPOSE of this paper to emphasize the possibility of asymmetric non-normal grade distributions, to present an example of this non-normality, to suggest some possible approaches to meet this problem, and to illustrate these approaches by a numerical example.

There has been considerable attention among educators in the past to student grade distributions. It has been repeatedly shown that for very large samples the normal distribution, although continuous, will often approximate the discrete numerical grade distribution. This result has been frequently misunderstood, and teachers have used the normal curve with some arbitrary standard deviation to say that a small class, no matter what the quality, should be ordered, the top, say, 7% to be given A, the next, say, 24% to be given B, etc.

The problems here involved are 1) confusing measurement and evaluation and 2) making unwarranted assumptions regarding the grade distribution. Some unwarranted assumptions include: a) students are placed in the class randomly, b) the numerical grade assigned to each student is assigned on the basis of degree of mastery of the material, i.e. an objective, numerical measure of performance, c) the distribution of performances of the students is a normal distribution, d) the assignment of letter grades by areas under the normal curve is made on the basis of some specified, arbitrary system, and e) the size of the class is adequate (a necessary but not sufficient condition) to justify this system.

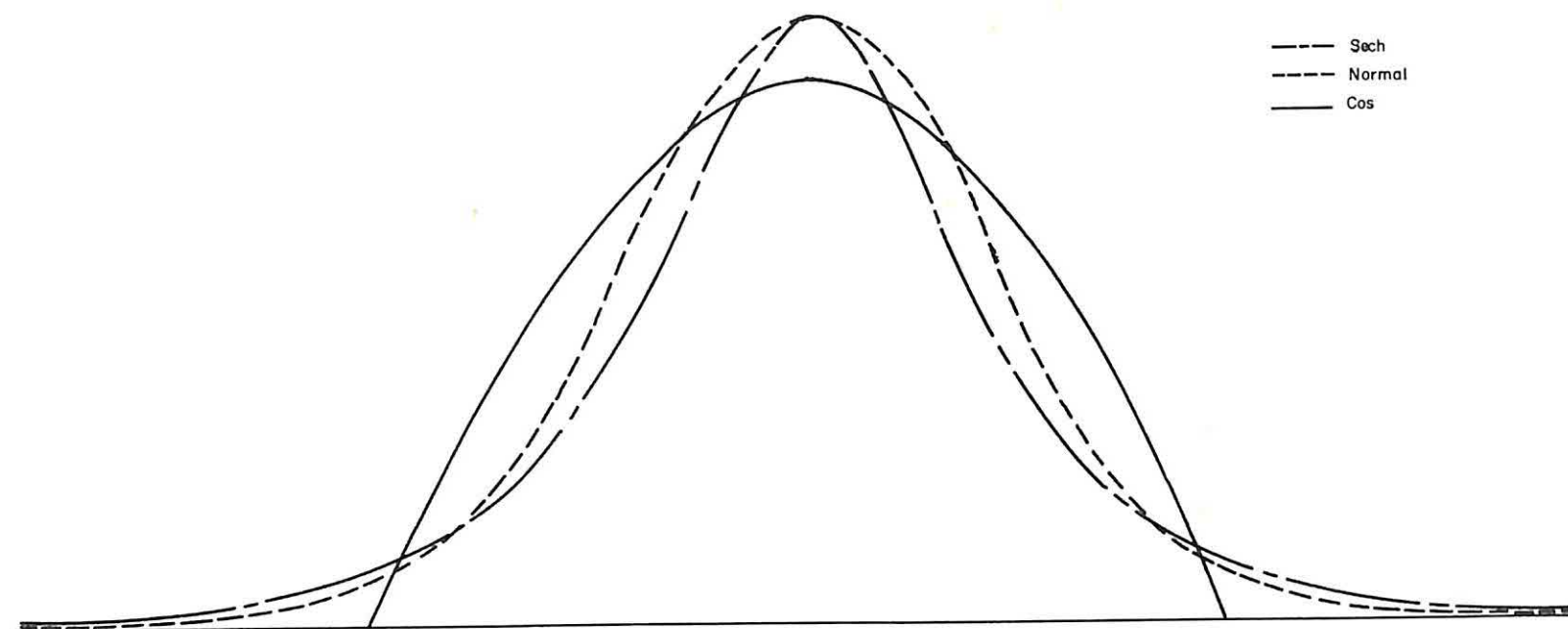
If classes are small, the teacher usually becomes personally acquainted with the ability of each student and the assignment of letter grades on an intuitive basis, considering students' background, effort, progress, etc. as well as objective performance, will often be more fair than making the necessary assumptions. When classes are large so that evaluation and measurement are necessarily identical, i.e. when the students are just names and grades are assigned solely upon performance measured by an objective criterion, then gross unfairness often results from making unwarranted assumptions.

This paper is concerned with the justification of

assumptions c) and d) when assumptions a), b), and e) are justified. Although the ability of the students may be normally distributed, it has been the author's repeated observation over several years of teaching that his students' standard, objective, numerical performance levels are not distributed symmetrically and therefore are not distributed normally. Many adequate students do less than adequate work because of illness, disinterest, interference of other activities, or other reasons unrelated to ability to perform. These students increase the frequency of low grades. However, the less-than-adequate students are unable, whatever the motive, to swell the high-grade bracket and thus do not balance the tendency to asymmetry. Thus, if assumption c) is unjustified, then obviously d) is violated. This implies that, when letter grades are repeatedly assigned with reference to a numerical grade distribution, the distribution should be investigated for asymmetry and non-normality. It is recommended that educators devise a "fair" method of assigning letter grades to students when the numerical grade distribution is non-normal.

If the numerical grade distribution should be non-normal, it is necessary to obtain an adequate approximation to the empirical distribution. Observation has suggested to the author a sum of curves, both of some bell-shaped, or at least concave downward, nature. It might be expected that the large curve expressing the greatest part of the area and having a mean not far from the distribution mean, might be normal. To this a smaller curve might be added, the nature of the curve to be indicated by its observed shape. Those shapes which might be expected to appear most regularly are another normal, a hyperbolic secant, the part of a cosine between 0 and $\pi/2$ radians. These three curves are depicted in Figure 1.

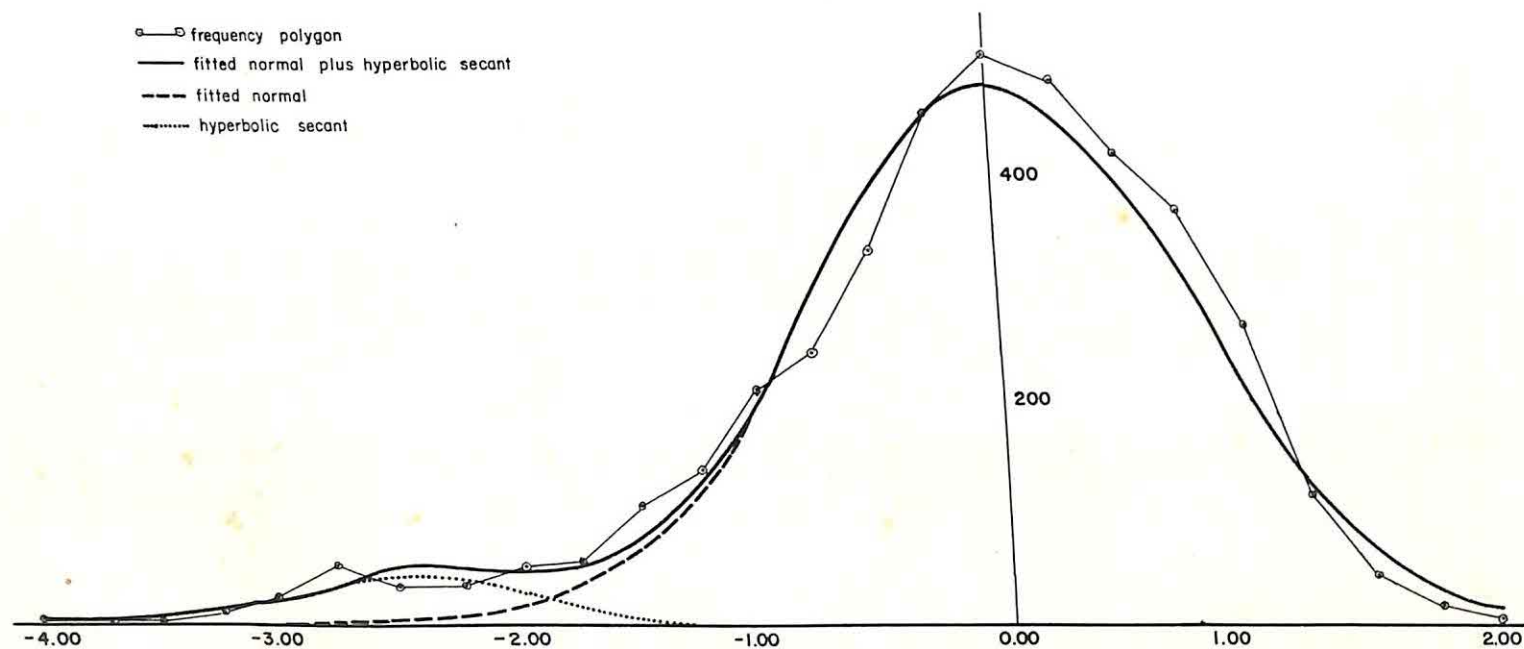
Sums of other curves may be appropriate in some cases, but not in the illustration of this paper. If, for example, the distribution should be not too far from normal, but have a heavy tail on one end, a part of a uniform curve, $y = k$; a linear curve, $y = ax + b$; or a parabolic curve, $y = ax^2 + bx + c$, might be appropriate. The distribution observed in



THE CURVES $y = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}$, $-\infty < x < \infty$; $y = \text{sech } x$, $-\infty < x < \infty$; AND, $y = \cos x$, $-\frac{\pi}{2} < x < \frac{\pi}{2}$;

SUPERIMPOSED UPON THE SAME AXIS WITH COINCIDENT MODES

FIGURE 1



FREQUENCY POLYGON FOR 4041 STUDENT GRADES WITH SUPERIMPOSED
FITTED CURVES

FIGURE 2

TABLE I

OBSERVED DATA AND FIT OF THIS DATA BY THE SUM OF NORMAL AND HYPERBOLIC SECANT CURVES

x	y_0	y_N	$y_0 - y_N$	prop. dev. 0 from N	y_S	$y_T =$ $y_N + y_S$	$y_0 - y_T$	prop. dev. 0 from T
2.00	3	16.69	13.69	.82	.00	16.69	13.69	.82
1.75	19	36.83	17.83	.48	.01	36.84	17.84	.48
1.50	48	73.45	25.45	.35	.01	73.46	25.46	.35
1.25	120	131.17	11.17	.09	.02	131.19	11.19	.09
1.00	268	211.13	56.87	.27	.03	211.16	56.84	.27
.75	376	305.40	70.60	.23	.07	305.47	70.53	.23
.50	427	398.20	28.80	.07	.12	398.32	28.68	.07
.25	489	465.47	23.53	.05	.21	465.68	23.32	.05
0.00	517	491.00	26.00	.05	.38	491.38	25.62	.05
-.25	456	465.47	9.47	.02	.67	466.14	10.14	.02
-.50	340	398.20	58.20	.15	1.19	399.39	59.39	.15
-.75	246	305.40	59.40	.19	2.12	307.52	61.52	.20
-1.00	214	211.13	2.87	.01	3.27	214.40	.40	.00
-1.25	138	131.17	6.83	.05	6.72	137.89	.11	.00
-1.50	106	73.45	32.55	.44	11.81	85.26	20.74	.24
-1.75	58	36.83	21.17	.57	20.41	57.24	.76	.01
-2.00	49	16.69	32.31	1.94	32.92	49.61	.61	.01
-2.25	33	6.78	26.32	3.88	45.24	52.02	19.02	.37
-2.50	35	2.50	32.50	13.00	46.74	49.24	14.24	.29
-2.75	59	.82	58.18	70.95	35.71	36.53	22.47	.62
-3.00	22	.25	21.75	87.00	22.48	22.73	.73	.03
-3.25	11	.06	10.94	182.33	13.26	13.32	2.32	.17
-3.50	3	.01	2.99	299.00	7.52	7.53	4.53	.60
-3.75	2	.00	2.00	∞	4.23	4.23	2.23	.53
-4.00	2	.00	2.00	∞	2.39	2.39	.39	.16

TABLE II
FIT OF DATA BY THE SUM OF TWO NORMAL CURVES

y_N	$y_T = y_N + y_{N'}$	$y_0 - y_T$	prop. dev. 0 from T
.00	16.69	13.69	.82
.00	36.83	17.84	.48
.00	73.45	25.46	.35
.00	131.17	11.19	.09
.00	211.13	56.87	.27
.00	305.40	70.60	.23
.00	398.20	28.80	.07
.01	465.48	23.52	.05
.04	491.04	25.96	.05
.15	465.62	9.62	.02
.53	398.73	58.73	.15
1.60	307.00	61.00	.20
4.14	215.27	1.27	.01
9.22	140.39	2.39	.02
17.48	90.93	16.93	.19
28.25	65.08	7.08	.11
39.30	55.99	6.99	.12
46.75	53.53	20.53	.38
47.90	50.40	15.40	.31
41.33	42.15	16.85	.40
30.60	30.85	8.85	.29
23.23	23.29	12.29	.53
10.61	10.61	7.61	.72
4.90	4.90	2.90	.59
1.95	1.95	.05	.03

TABLE III
FIT OF DATA BY THE SUM OF NORMAL AND COSINE CURVES

y_C	$y_T = y_N + y_C$	$y_0 - y_T$	prop. dev. 0 from T
.00	16.69	13.69	.82
.00	36.83	17.83	.48
.00	73.45	25.45	.35
.00	131.17	11.17	.09
.00	211.13	56.87	.27
.00	305.40	70.60	.23
.00	398.20	28.80	.07
.00	465.47	23.53	.05
.00	491.00	26.00	.05
.00	465.47	9.47	.02
.00	398.20	58.20	.15
2.56	307.96	61.96	.20
9.13	220.26	6.26	.03
15.32	146.49	8.49	.06
20.69	94.14	11.86	.13
25.58	62.41	4.41	.07
28.08	44.77	4.23	.09
29.24	36.02	3.02	.08
29.86	32.36	2.64	.08
28.53	29.35	29.65	1.01
25.74	25.99	3.99	.15
22.54	22.60	11.60	.51
16.47	16.48	13.48	.82
10.42	10.42	8.42	.81
3.92	3.92	1.92	.49

any special case will suggest the appropriate curve to use.

Numerical Example

A tabulation of 4041 numerical grades from freshmen and sophomore mathematics courses was made from files of the University of Hawaii mathematics faculty. Letter grades were not considered. In order to adjust for the obvious disparity of varying standards of professors, each professor's grade distribution was standardized to have zero mean and unit standard deviation. The grades were recorded as fourths of standard deviations from the mean. The distributions were then pooled. The frequency polygon of this sample of size 4041 is depicted in Figure 2.

Clearly there is a large curve which may be approximated by the normal enclosing the bulk of the area, but there is also a "bump" on the negative tail associated with low, or "F" grades.

The normal curve was fitted by trial and error; standard methods were inappropriate since the distribution mean and variance would not coincide with the fitted normal's mean and variance. This normal fit is given as a dotted line in Figure 2. Table I gives, for each abscissa value, x , the corresponding ordinate value observed from the data, y_0 , the corresponding ordinate value expected from the normal curve fit, y_N , the difference, $y_0 - y_N$, and the proportion of deviation of the observed curve from the theoretical curve. Inspection shows that this proportion becomes large toward the negative tail of the distribution, indicating a remarkably poor fit

in that region.

Subsequently, the three curves (normal, hyperbolic secant, cosine) suggested as possibly appropriate were added to the original normal. From observation it was thought that the hyperbolic secant would be the most appropriate curve; the others were included for illustration. The remainder of Table I is composed of the fitted hyperbolic secant ordinate corresponding to each abscissa value, y_S , the corresponding sums of the two curves, y_T (ordinate values expected from the total fitted curve), the difference of the observed and expected values, $y_0 - y_T$, and this difference expressed as a proportion deviation. The last column shows that the deviation of the expected values from the observed values is reasonably small. A comparison of the two deviation columns shows that, while the normal fit is distinctly inadequate, the fit by the sum of curves is satisfactory. The final curve is given by the equation:

$$y = \frac{944.6}{\sqrt{2\pi}(.768)} \exp \left\{ -\frac{1}{2} \frac{x^2}{.59} \right\} + 48 \operatorname{sech} 2.31(2.4 + x)$$

Tables II and III repeat the last four columns of Table I (the first five are identical), but give a normal fit and a cosine fit replacing the hyperbolic secant fit. From the deviation columns it can be observed that these latter two fits are nearly as accurate as the first fit, and that they improve the results remarkably over the initial normal fit. The complete equations involving the second normal and the cosine curves (in radians) respectively are:

$$y = \frac{944.6}{\sqrt{2\pi}(.768)} \exp \left\{ -\frac{1}{2} \frac{x^2}{.59} \right\} + \frac{76.0}{\sqrt{2\pi}(.632)} \exp \left\{ -\frac{1}{2} \frac{(x + 2.4)^2}{.4} \right\}$$

and

$$y = \frac{944.6}{\sqrt{2\pi}(.768)} \exp \left\{ -\frac{1}{2} \frac{x^2}{.59} \right\} + 30 \cos .9(x + 2.4) .$$

SOCIOECONOMIC STATUS, URBANISM AND ACADEMIC PERFORMANCE IN COLLEGE

NORMAN F. WASHBURN*
Alexandria, Virginia

DEANS, ADMISSION officers and counselors, in their attempts to evaluate prospective students, have become greatly concerned with the factors that make for scholastic success in college. Attempts to construct valid predictive indexes have usually concentrated on high school performance, various test scores, and personality factors such as those which come to light during brief interviews and in letters of recommendation from teachers, principals and alumnae. Social factors are rarely openly used in evaluating applicants for admission to college, not only because some of them are hard to measure, but also because most colleges wish to avoid any accusation that such things as social class, race, nativity or religion, are factors in their selection procedures.

Consequently, little is known about the relationship of sociological variables to academic success even though sociological theory leads one to expect that such relationships exist. For instance, differential socioeconomic status implies differential cultural opportunities and variation in the focus of interests of college students, and it seems reasonable to believe that both cultural opportunity and interests are factors in performance in the classroom. Urbanism is another factor which is theoretically related to academic performance, because such educationally important factors as attitudes, and levels of aspiration, as well as cultural opportunities are different among urbanites than they are among rural people.¹

This paper is a report of an attempt to investigate the relationships of socioeconomic status and urbanism to academic performance. Specifically, the hypothesis was that academic performance would be positively and significantly correlated with the socioeconomic status of families of college students and with the degree of urbanism of the communities in which they had grown up.

Samples were drawn from two widely different college populations. One college is a state-supported institution located in a small city (population approximately 7,000) in a relatively sparsely set-

tled part of the Southwest. The other college is a privately endowed institution located in a major Northeastern metropolis. The use of these two colleges was indicated by the desire to avoid findings which might be unique to a particular college or region, and also by the desire to have as wide a range as possible both in socioeconomic status and in urbanism. Each of the samples consisted of 100 randomly selected entering freshmen. All of the students were males.

The research design called for the development of measures which would be applicable to each student and which would quantify his academic performance, the socioeconomic status of his family, and the degree of urbanism of each of the places he had lived in while growing up. If the measures of socioeconomic status and urbanism were found to be positively correlated with the measure of academic success in each of the two samples, then the hypothesis would be confirmed.

The measure of academic performance in this study was the mean of the grade point averages earned by the student during his first two semesters in college. The grade point average is computed in such a manner that 4.00 would indicate a straight "A" record, while 0.00 would indicate a straight "F" record. The distributions of mean semester grade point averages for the two samples are summarized in Table I.

It is somewhat more difficult to arrive at a valid and accurate measure of the socioeconomic statuses of the students' families. This is the case because most studies of the factors comprising such status have been restricted to single communities, and there is every reason to believe that what is true in one community may well not be true in another. It seems particularly likely that the specifics of the status structure of the Southwest are different from those in the Northeast. On the other hand, there is evidence that in both regions education and occupation have a great deal to do with a family's social standing, and it seemed that a socioeconomic status scale based on these two factors

* Consulting Psychologist, 51 Northdown Road.

1. See Washburne, *Attitudes and Responses of College Students as Related to their Residence in Urban Communities and to their Socioeconomic Status*, University Microfilms, Inc., 1953.

TABLE I
SUMMARY OF DISTRIBUTIONS OF MEAN SEMESTER
GRADE POINT AVERAGES

Southwestern College Sample	Northeastern College Sample
Range: 0.75 to 4.00	Range: 0.19 to 3.96
Mean: 2.17	Mean: 2.05
Standard Deviation: 0.68	Standard Deviation: 0.72

TABLE II
SUMMARY OF DISTRIBUTIONS OF SOCIOECONOMIC STATUS SCORES

Southwestern College Sample	Northeastern College Sample
Range: 2 to 10	Range: 4 to 10
Mean: 6.11	Mean 7.06
Standard Deviation: 1.9	Standard Deviation: 1.7

could be used, provided it was adjusted for regional differences.

In developing the educational aspect of the socioeconomic scale, the basic data used consisted of statements of the number of years of formal education completed by both the father and the mother of each of the students. When the resultant frequency distributions were inspected, it became apparent that the parents of both sets of students tended to have finished their schooling at one of the following five points: the end of the sixth grade, the end of the ninth grade, the end of the 12th grade, the end of college, or after completing some graduate or professional work. Therefore, it seemed logical to convert years of education to the following five point scale:

0 - 6 years	- 1
7 - 9 years	- 2
10 - 12 years	- 3
13 - 16 years	- 4
17 + years	- 5

The educational factor in the socioeconomic status score for each student was arrived at by computing the scores corresponding to the years of education completed by his mother and by his father, and then averaging the two scores together. The use of the education of both parents for this score seems justified in that the backgrounds of each contribute to the tenor of home life, and hence to the status orientation of the family.

The basic data of the occupational aspects were the present occupations of the parents of the students. Where there was more than one job, the highest occupation was used. The relative weights for the occupations were different for each sample, because there is a regional difference in both the relative prestige of similar occupations and in the types of jobs the parents held. For instance, many of the parents of the Southwesterners were engaged in agriculture, whereas only one of the Northeastern parents was. The weights were arrived at by listing the occupation of the parents of the students in the sample among many other jobs of all types. Then a new sample of 100 students was drawn and this sample was asked to rate each occupation on a five point scale ranging from "excellent" to "poor" in general social standing.² Then after compensating for the halo effect, the occupations of the parents were weighted from 1 for poor occupations to 5 for excellent ones, according to the judgment of the jurors. This process was repeated at each school so that regional differences could be compensated. Incidentally, some of the regional differences were revealing. For instance, the area around the Southwestern college was in the midst of an oil boom, and the students there rated "petroleum engineer" higher than "chemical engineer," whereas the Northeastern students quite logically

rated them both the same. Proximity to one's own politicians seems to lower one's esteem for them, for the Northeasterners rated "Mayor of a small town" higher than did the Southwesterners, whereas the Southwesterners thought more of a "Mayor of a large city" than did their urban Yankee counterparts.

The final socioeconomic status score for each student was computed by adding the educational factor, numerically expressed, to the weight given the highest occupation presently held by either of his parents. The resultant measure could vary from a low status score of 2 to a high of 10. The distributions of socioeconomic status scores for the two samples were normal and are summarized in Table II.

Finally, a measure was needed which could differentiate among the students in terms of the urbanism of their backgrounds. The simplest way would have been to have classified them in terms of whether their home towns were listed by the Bureau of the Census as being urban or rural. However, there are two major objections to such a procedure. In the first place, any dichotomization of communities is bound to hide great within-class variation. Consider, for instance, the great difference in cultural opportunities that exists between Hope, Arkansas, and New York City, yet each of these places is urban according to the census definition.

For these reasons it seemed necessary to construct an index which would measure each student's experience in communities of varying degrees of urbanism throughout their lives.

The problem of constructing such an index was essentially one of assigning degrees of urbanism to each place where the student had lived during his formative years. To begin with, a place in a large town could be regarded as more urban than one in a small town. However, this cannot be the only consideration, for obviously a place in a community of 2,500 a few miles away from New York City could hardly be called less urban than a place in a community of, say, 7,500--three times as populous--which is in the middle of the country, many, many miles from the nearest larger town. Therefore, in assigning degrees of urbanism to communities, consideration must not only be given to the size of the place itself, but also to the proximity of larger urban aggregates.

Each respondent completed a questionnaire which yielded the following information about each place of residence in which he had lived since the time he first entered the first grade of school:

1. The name of the community in which the place was located, or the name of the nearest community.
2. The inclusive dates of his residence there.
3. Whether or not the place was inside the town limits, and if not, how far the place of residence

2. This method is the same as that developed by North and Hatt. See: C. C. North and P. K. Hatt, "Jobs and Occupations, a Popular Evaluation," *Opinion News*, September 1, 1947, pp. 3 - 13.

was from the town.

Then the 1950 population of the community was ascertained from the Census of the Population. If the community was a part of a metropolitan district, the population of the district was used, rather than the population of the community.

Then the following distances in highway miles were ascertained for each community from a standard highway map:

1. The distance to the nearest town of 250 population or more.
2. The distance to the nearest town of 500 population or more.
3. The distance to the nearest town of 1,000 population or more.
4. The distance to the nearest town of 2,500 population or more.
5. The distance to the nearest city of 5,000 population or more.
6. The distance to the nearest city of 10,000 population or more.
7. The distance to the nearest city of 25,000 population or more.
8. The distance to the nearest city or metropolitan district of 50,000 or more.
9. The distance to the nearest city or metropolitan district of 100,000 or more.
10. The distance to the nearest city or metropolitan district of 250,000 or more.
11. The distance to the nearest city or metropolitan district of 500,000 or more.
12. The distance to the nearest city or metropolitan district of 1,000,000 or more.
13. The distance to the nearest city or metropolitan district of 2,500,000 or more.
14. The distance to the nearest city or metropolitan district of 5,000,000 or more.
15. The distance to the nearest city or metropolitan district of 10,000,000 or more.

These fifteen distances were expressed for each community in terms of the following scale:

- 0 to 6 miles - 5
- 7 to 12 miles - 4
- 13 to 24 miles - 3
- 25 to 49 miles - 2
- 50 to 99 miles - 1
- 100 to more miles - 0

After the distance had been so expressed, the resultant points were totaled. This sum was the urbanism score for the community.

In order to clarify the scheme, two examples are offered. The first is illustrated in Figure 1 which should be referred to in the discussion to follow. Emerson is a little community in Arkansas. In 1950 the Census reported 523 residents there. Emerson is 12 miles from Magnolia, Arkansas, which is the nearest larger urban aggregate. The urbanism score for Emerson works out as follows: Emerson has itself more than 250 population, so that it receives 5 points for the first of the fifteen measures. It also has a population of more than 500, so it receives another 5 for the second measure.

The nearest community of 1,000 or more population is Magnolia, which had in 1950 a little more than 6,300 residents. So Emerson, twelve miles from Magnolia, gets 4 points for the third, fourth and fifth measures. Eldorado, a city of between 10,000 and 25,000 population is within 49 miles of Emerson, so two more points are added for the sixth measure. Texarkana, which has more than 25,000 population, is the nearest city of that size and is more than 50, but less than 100 miles away, so Emerson receives a point for the seventh measure. Shreveport is the nearest city of 50,000 or more population, and it is more than fifty, but less than one hundred miles away, so Emerson gets one more point for the eighth measure. Shreveport has more than 100,000 residents, so Emerson gets a point in the ninth measure. There are no cities of 250,000 or more population within 100 miles, so Emerson gets no points for the eighth through the fifteenth measures, for a total score of 27.

In contrast, Scarsdale, New York, is inside a Metropolitan district which had a population of more than 10,000,000 people in 1950. Therefore, Scarsdale receives 5 points for each of the 15 measures, for a total score of 75.

Some places of residence are, of course, not inside communities. A few of the respondents in the Southwestern sample had lived on isolated farms as much as 25 miles from the nearest town of any size at all. Therefore, in order to compute the urbanism score for each place of residence, a sixteenth measure was added; namely, the distance for the place of residence from the community to which the fifteenth measures had been applied. Thus, a residence on a farm seven miles away from Emerson, the nearest community, would receive an additional four points to bring the total score for that place of residence to 31. If he had lived inside the town limits, the place would have received five rather than four extra points.

After the score of each place of residence for each student had been computed, each score was multiplied by the number of years he had lived there since entering grammar school. The resultant products were for each respondent totaled and then divided by number of years he had lived in all places since entering school. The result is the score of urbanism of the residence history of the student, and is the measure of urbanism used in this study.

By way of example, if a nineteen-year-old student had lived the first twelve years of his life on that farm seven miles from Emerson, entering school there at the age of six, and then had moved to Scarsdale at the age of twelve, and had lived there ever since, his score of urbanism would be computed as shown at the top of the next page.

The urbanism scale is based on assumptions concerning the relationships of size of place of residence, degree of that place's isolation, and its proximity to urban aggregates, to the nature of the cultural milieu in which children grow up. The scale is applicable to individuals and has been de-

URBANISM SCALE EMERSON, ARK

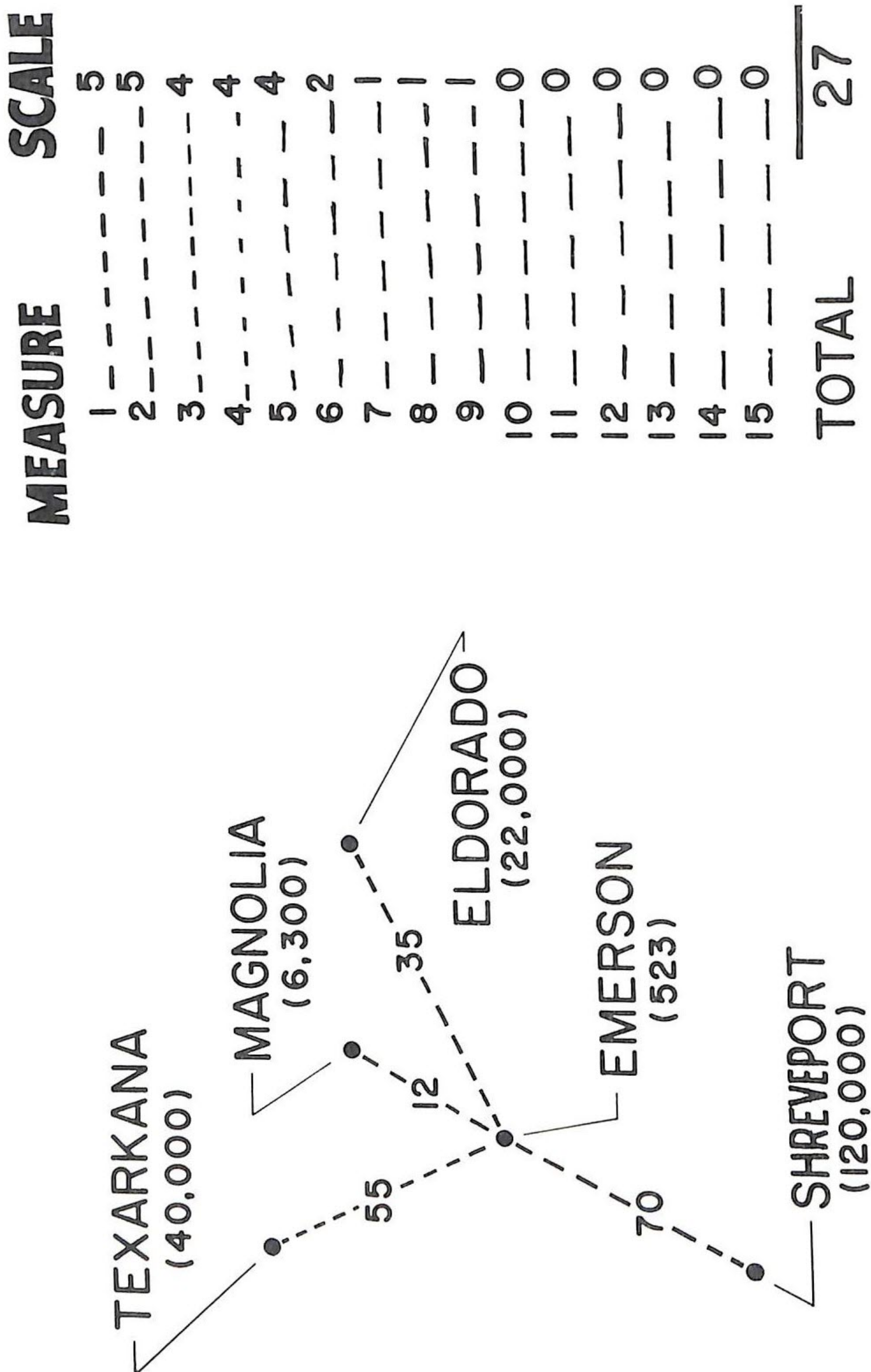


FIGURE 1.

TABLE III
SUMMARY OF DISTRIBUTIONS OF URBANISM SCORES

Southwestern College Sample	Northeastern College Sample
Range: 10 to 53	Range: 26 to 80
Mean: 28.5	Mean: 61.2
Standard Deviation: 9.6	Standard Deviation: 7.8

TABLE IV
COEFFICIENTS OF TOTAL CORRELATION OF SOCIOECONOMIC STATUS
SCORES AND URBANISM SCORES WITH THE MEAN
SEMESTER GRADE POINT AVERAGES

	Southwestern College Sample	Northeastern College Sample
Socioeconomic status scores with Mean Semester Grades, $r =$	-.05	-.06
Urbanism of residence scores with Mean Semester Grades, $r =$	+.31	+.18

NORTHEASTERN COLLEGE

GRADES

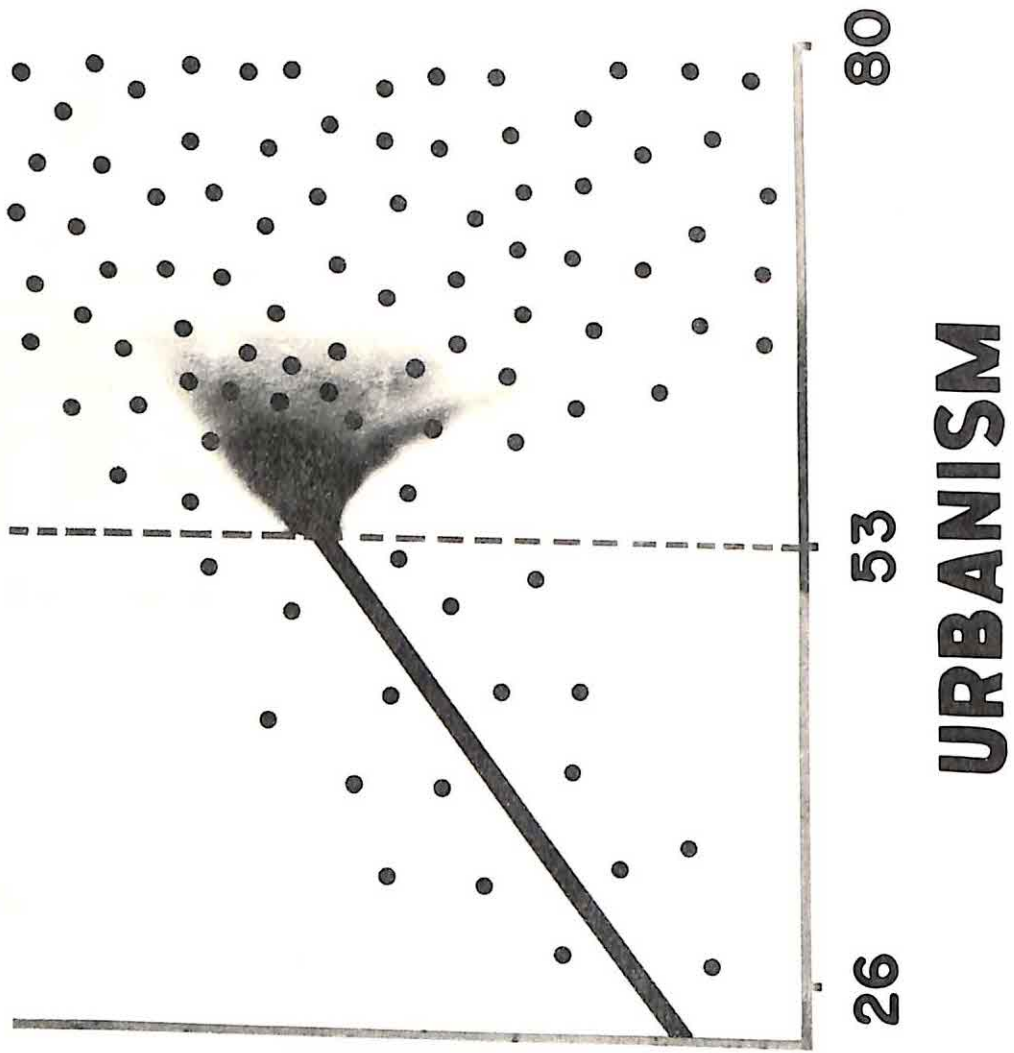


FIGURE 2.

(Score of place of Residence) \times (Years since entering school) = Product

Farm near Emerson:	31	\times	6	=	186
Inside Scarsdale:	80	\times	$\frac{7}{13}$	=	$\frac{560}{13}$
Totals:					746

Score of urbanism of residence history = $\frac{746}{13} = 57$.

monstrated to be predictive of attitude areas such as religious tolerance, traditionalism, and "familism" among college students.³

The distributions of urbanism scores of the two samples are summarized in Table III.

After it was determined that the independent variables (the urbanism scores and the socioeconomic status scores) were not significantly correlated, coefficients of correlation were computed for each of them with the mean semester grade point averages of the students in each of the two samples. The findings are presented in Table IV.

Socioeconomic status was not significantly related to academic performance for either of the samples, at least as these variables were measured here. Perhaps the part played by socioeconomic status as regards higher education is limited to the determination of opportunity to attend school, and the development of levels of aspiration such that the student is motivated to attend when the opportunity is offered. At any rate, evidence here indicates that once the student gets to college, the socioeconomic status of his family has nothing to do with his academic performance.

Urbanism was correlated positively and significantly with academic success for the students at the Southwestern college. However, for the Northeastern group this relationship, while positive, is so low in magnitude that there is more than a five percent probability that it occurred by chance because of the sampling procedure. The possibility that the relationship between urbanism and grade point average was a phenomenon related to the rapid social change currently occurring in the Southwest

was seriously contemplated, until examination of the scatter plot for the Northeastern group brought to light an interesting fact (See Figure 2.) The plot revealed that for the students on the lower end of the urbanism continuum, the positive regression of grades seem to exist, while for the students on the upper end of the continuum, the relationship completely broke down, and for the very urban students, variation in academic performance had nothing to do with variation in urbanism.

To test this visual impression, a correlation coefficient was computed for the urbanism scores and the mean semester grade point averages of the 18 Northeastern students who had urbanism scores of 53 or less--the same upper limit as the Southwestern sample. In this case, $r = +.37$, approximately the same magnitude as for the Southwesterners. For scores above 53, there was no correlation whatever between urbanism and academic performance.

Thus, it can be seen that for both samples, the more urban the residence background of the student, the better his academic performance is likely to be--up to a point. That point seems to be the 500,000 population mark, for students who have lived most of their lives in or near major metropolitan districts vary from one extreme to the other in academic performance.

It would be interesting to investigate the social correlates of academic performance among a population made up exclusively of students with metropolitan backgrounds. One thing is apparent, though. The evidence here indicates that socioeconomic status has little, if anything, to do with it.

3. Washburne, op. cit.

THE RELATIONSHIP OF SELECTED MEASURES TO SPELLING ACHIEVEMENT AT THE FOURTH AND EIGHTH GRADE LEVELS

I. E. AARON
University of Georgia

Purpose of the Study

This study attempted to determine the relationship of selected skills and abilities to spelling achievement among fourth and eighth grade children. It was a follow-up of a larger, more inclusive study carried out at the fourth grade level.^{1*}

In day-to-day discussion among teachers, many differences of opinion are voiced about the relationship of phonics and syllabication to spelling achievement. Professional articles on the teaching of spelling are frequently in disagreement on this point, and the extent to which spelling texts make use of phonics and syllabication in suggested teaching exercises varies because of differences of opinion on the part of authors. For these reasons, further study of the relationship of these skills to spelling achievement appears to be warranted.

The Previous Study

Earlier research, involving 188 fourth grade children from one of Georgia's approximately 200 school districts, studied the relative contributions of nine auditory-visual discrimination measures and intelligence in estimating five spelling measures. The auditory-visual discrimination measures were visual analysis of words (including syllabication),² recognition of beginning and ending sounds in words,³ knowledge of sound values of common word elements,⁴ word discrimination ability,⁵ spelling of phonetic syllables, and four measures of binaural acuity.⁶ The spelling tests contained words divided into purely phonetic words, almost phonetic words, phonetic words (a combination of the previous two), partially phonetic words, and non-phonetic words.

Of the measures, only three were found to be of value in estimating spelling achievement. These were visual analysis of words, recognition of beginning and ending sounds in words, and spelling of phonetic syllables. Intelligence, though positively correlated with the spelling measures, did not con-

tribute significantly to the multiple regression equations for predicting spelling achievement when in combination with the other measures.

The study reported here makes use of the measure, spelling of phonetic syllables, that contributed the most to predicting spelling achievement and the same intelligence test of the earlier investigation. A newly constructed test of syllabication was used. Syllabication was a part of the earlier test on visual analysis of words. In general, the statistical treatment of the previous research was the same as that of the present study. The investigation reported here utilized both fourth and eighth grade pupils to determine if the relationship found at the fourth grade level also held at the eighth grade level.

Tests Used

A 60-word, list-dictation spelling test composed by the writer was used to measure spelling achievement. It contained words of varying degrees of phonetic composition, ranging from "purely phonetic" to "non-phonetic." "Purely phonetic" words were those whose pronunciation spellings used exactly the same letters in the precise order of the correct spelling of the words, provided, further, that all phonemes in the words were commonly attributed to the letters used to represent these in the pronunciation spellings. Words in the test varied from the purely phonetic to those that were non-phonetic in at least one component. Words were selected from the three 60-word tests used in the previous study. Twenty words were selected randomly from each of the three tests. Words were based upon those contained in Ashbaugh's list.⁷ By use of the Hoyt Analysis of Variance method⁸ for computing test reliability, test reliabilities were found to be 0.97 and 0.96 for the fourth and eighth grades respectively. These were based upon the scores of the 57 fourth grade pupils and the 54 eighth grade pupils in the largest school participating in the study. The spelling of phonetic syllables test was com-

*Footnotes will be found at the end of the article.

posed of fifteen nonsense monosyllables and fifteen nonsense plurisyllables. It contained such nonsense words as "whong" and "checkzub." Ten of the monosyllables and eight of the plurisyllables were taken from an oral examination used by Gates and Russell,⁹ and the others were prepared by the writer in keeping with the Gates and Russell general design. Whereas the Gates and Russell test presented the words orally and had the subjects answer orally, in this study the syllables were presented in a group situation and pupils wrote their responses. Each syllable was presented twice. Scoring was based upon a scale on which the syllable could be credited with two, one, or no points. Full credit was given to those spellings nearest to the actual sound of the syllable whereas partial credit was given when the spellings were correct but not the spelling which would most logically be associated with the sound of the syllable. This test was included because it was believed that such a test would be an almost pure measure of a pupil's ability to transmit an oral sound into its symbolic equivalent. The Hoyt method revealed reliability coefficients of 0.97 for the fourth grade and 0.96 for the eighth grade. Again the scores of the pupils in the largest school participating in the study were used for reliability computations.

The non-language section of the California Intelligence Test was used as a measure of intelligence. This test was selected because the reading factor was minimized, and pupils who were disabled in reading were penalized less with this type of test than with other available intelligence tests.

The syllabication test, the only new test used in this study, consisted of 25 nonsense words that the pupil was asked to divide into syllables. They involved four syllabication situations: 1) single consonant between two vowels, 2) double consonants between two vowels, 3) words ending in "le" following a consonant, and 4) the suffix "ed" following "d" or "t". The pupil's score was the number of words he got correct. A mimeographed sheet containing the words was passed out to each child, and pupils were asked to draw a line between the syllables in the words. The test contained such nonsense words as "tumil" and "siblog." The Hoyt method, utilizing all pupils in the study, revealed a test reliability of 0.54 at both the fourth and eighth grade levels.

The populations of this study consisted of all of the white, fourth and eighth grade pupils in one of Georgia's approximately 200 school districts. This was not the same school system involved in the first study. The fourth grade group included 193 pupils, and the eighth grade group consisted of 174 pupils. Children came from five different schools and from a variety of home backgrounds.

The hypotheses tested in this study were as follows:

1) Ability in spelling of phonetic syllables does not contribute significantly in predicting spelling achievement.

2) Syllabication ability does not contribute significantly in predicting spelling achievement.

3) Intelligence does not contribute significantly in predicting spelling achievement.

Statistical Procedures

The statistical tool appropriate for such an analysis as that of this study is multiple regression. Several of the distributions of the variables were found to be non-normal when tested by means of inspection of plots on probability paper. All variables were transformed by means of probits, and all computations, with the exception of test reliabilities, were based upon transformed scores. Computations for fourth and eighth grades were treated separately throughout the study.

Product moment correlation coefficients were computed between each of the four measures and every other measure. The significance of each was tested by means of a special table of *t* values designed for this purpose.¹⁰ Table I presents the obtained correlation coefficients at both fourth and eighth grades. It will be noted that the largest correlation at both grade levels is that between spelling and spelling of phonetic syllables. All coefficients of correlation were found to be significant at the .01 level.

The investigation of standard partial regression weights was designed to discover the relative contributions of each of three independent variates in predicting spelling achievement. By use of standard partial regression coefficients, it was possible to determine the relative importance of each of the independent measures (spelling of phonetic syllables, syllabication, and intelligence) in predicting the dependent measure (spelling).

Throughout the study, fourth and eighth grade data were treated separately though they are presented jointly in the tables included in this report.

At both grade levels, analysis was begun by computing standard partial regression weights for the three independent measures in predicting the dependent measure. Standard partial regression coefficients obtained were tested with *t* tests.¹¹ If the probability of getting coefficients of each magnitude by chance in repeated sampling was 1 in 100 or less, the coefficients were considered to be significantly different from zero. If the probability of getting weights of such size or greater by chance in repeated sampling was between 1 and 5 in 100, the hypothesis regarding significance remained in doubt. However, if sizes of coefficients were such that they could have arisen by chance in repeated sampling more than 5 times per 100 samples, they were considered not significantly different from zero. Those variates not significantly larger than zero were eliminated and the regression weights recomputed. Those variates in doubt (between .01 and .05) were retained. For each set of regression weights, a multiple correlation coefficient was computed in order to determine the predictive ef-

TABLE I
PRODUCT MOMENT CORRELATION COEFFICIENTS ON FOUR MEASURES, 193
FOURTH GRADE PUPILS AND 174 EIGHTH GRADE PUPILS

Measures Correlated	Fourth Grade	Eighth Grade
Spelling and Spelling of Phonetic Syllables	0.8449*	0.6177*
Spelling and Syllabication	0.4193*	0.3954*
Spelling and Intelligence	0.4361*	0.4607*
Spelling of Phonetic Syllables and Syllabication	0.3985*	0.2856*
Spelling of Phonetic Syllables and Intelligence	0.4380*	0.5057*
Syllabication and Intelligence	0.2300*	0.3898*

* Significant at the .01 level

TABLE II
STANDARD PARTIAL REGRESSION WEIGHTS FOR PREDICTING SPELLING
ACHIEVEMENT FROM THREE INDEPENDENT MEASURES,
FOURTH AND EIGHTH GRADES

Variable	Fourth Grade		Eighth Grade	
	Weight	t Value	Weight	t Value
Spelling of Phonetic Syllables	0.7748	17.2178*	0.4928	7.3224*
Syllabication	0.0931	2.2326**	0.2030	3.2222*
Intelligence	0.0753	1.7759	0.1076	1.5371

* Significant at .01 level ** Significant at .05 level

TABLE III
STANDARD PARTIAL REGRESSION WEIGHTS FOR PREDICTING SPELLING
ACHIEVEMENT FROM TWO INDEPENDENT MEASURES, FOURTH AND
EIGHTH GRADES, WITH INTELLIGENCE MEASURE OMITTED

Variable	Fourth Grade		Eighth Grade	
	Weight	t Value	Weight	t Value
Spelling of Phonetic Syllables	0.8058	19.3237*	0.5390	8.9091*
Syllabication	0.0981	2.3525**	0.2318	3.8314*
* Significant at .01 level		** Significant at .05 level		

TABLE IV
TESTS OF SIGNIFICANCE OF DIFFERENCES BETWEEN MULTIPLE CORRELATIONS

Grade	R for 3 Variates	R for 2 Variates	F Value	Hypothesis
Fourth	.8523	.8496	3.2142*	Accepted
Eighth	.6590	.6516	2.9393*	Accepted
* Probability greater than .05				

iciency of the equations. The analysis of variance technique was used to test the significance of the multiple correlations.¹² As a final step when variables were dropped, tests of significance of differences were made between the multiple correlations involved. Analysis of variance tests were again used.¹³

Table II shows the results of the initial computations of standard partial regression weights at both fourth and eighth grade levels. The standard partial regression weight of spelling of phonetic syllables in predicting spelling achievement was significantly different from zero for both fourth and eighth grade levels. These standard partial regression weights were by far the largest of the weights computed at each grade level. The standard partial regression weight for syllabication in predicting spelling achievement was significantly different from zero at the .05 level for the fourth grade and at the .01 level for the eighth grade. The weight for intelligence in predicting spelling achievement was not significantly different from zero for either the fourth or eighth grade groups.

To summarize, in the fourth grade, the standard partial regression weight of spelling of phonetic syllables was significantly different from zero at the .01 level and the syllabication weight at the .05 level. The standard partial regression weights of both of these measures in the eighth grade were significantly different from zero at the .01 level. Intelligence, in both grade groups, did not contribute significantly to the prediction of spelling achievement since the standard partial regression weight for intelligence was not significantly different from zero.

The next computational step was to drop intelligence from the independent measures and recompute the standard partial regression weights of the two remaining independent measures. This was undertaken because intelligence did not contribute significantly to the prediction of spelling achievement when in combination with the two remaining variates. Table III shows the resulting standard partial regression weights for both the fourth and the eighth grades. Again, the standard partial regression weight of spelling of phonetic syllables was significant at the .01 level in both fourth and eighth grade groups. As in the previous table, syllabication was significantly different from zero in the fourth grade group at the .05 level and in the eighth grade group at the .01 level.

The previously computed multiple correlation coefficient, using all three independent measures, was next compared with the multiple correlation coefficient computed from the two remaining independent measures at fourth and eighth grades separately. *F* tests were again used for making these comparisons. They are summarized in Table IV. As may be noted, the multiple correlation coefficients did not differ significantly, indicating that the dropping of intelligence did not affect significantly the prediction of the spelling measure when it was used

in combination with the other two independent measures.

Though the standard partial regression weight for the fourth grade group was of doubtful significance, it was not dropped from the equation. It was significantly different from zero at the .01 level in the eighth grade group.

Summary and Conclusions

This investigation, involving 193 fourth grade pupils and 174 eighth grade pupils, studied the relationship of three selected skills and abilities to spelling achievement. By use of multiple regression, the relative importance of 1) spelling of phonetic syllables, 2) syllabication, and 3) intelligence in predicting spelling achievement was studied. Fourth and eighth grade data were treated separately.

This study dealt with selected populations of fourth and eighth grade children. Generalizations from this study should be limited to the particular populations utilized in the study or other populations of similar composition.

From the statistical treatment of the data, the following conclusions may be drawn:

1) The hypothesis that spelling of phonetic syllables ability does not contribute significantly in predicting spelling achievement was rejected at both fourth and eighth grade levels.

2) The hypothesis that syllabication ability does not contribute significantly in predicting spelling achievement remained in doubt at the fourth grade level and was rejected at the eighth grade level.

3) The results failed to refute the hypothesis that intelligence does not contribute significantly in predicting spelling achievement when in combination with the other two measures at both fourth and eighth grade levels.

4) By far the greatest contribution to the estimation of spelling achievement at both fourth and eighth grade levels came from the spelling of phonetic syllables test.

The fact that the spelling of phonetic syllables test is a heavy contributor to the prediction of spelling success in this research gives added support that phonics skills are related to spelling achievement. The relationship noted is not necessarily causal. Further research is needed to determine whether or not the relationship found is causal.

FOOTNOTES

1. Aaron, Ira E., The Relationship of Auditory-Visual Discrimination to Spelling Ability. Unpublished Ph.D. thesis, University of Minnesota, 1954.
2. This test was part of an early experimental form of the Bond, Clymer, Hoyt Silent Reading Diagnostic Test.
3. Ibid.
4. Ibid.

5. This was a 50-item test designed to measure the pupil's ability to discriminate between similar sounding words. The test was administered by use of a tape recorder so that pronunciation would be uniform from group to group.
6. These measures consisted of high, medium, and low frequency and total scale scores taken from full scale audiograms for each pupil.
7. Ashbaugh, Ernest J., The Iowa Spelling Scales. College of Education, University of Iowa, Journal of Educational Research Monographs, No. 3. 1922. 144 pp.
8. Hoyt, Cyril J., "Note on a Simplified Method of Computing Test Reliability." Educational and Psychological Measurement. 1: 93-95, 1941.
9. Gates, Arthur I. and Russell, David H., Diagnosics and Remedial Spelling Manual. Bureau of Publications, Teachers College, Columbia University, New York. 1940 (revised).
10. Garrett, Henry E., Statistics in Psychology and Education. Fourth Edition. Longmans, Green and Company, Inc., New York. 1953. pp. 437-439.
11. Johnson, Palmer O., Statistical Methods in Research. Prentice-Hall, Inc., New York, 1949, pp. 338-339
12. Goulden, C. H., Methods of Statistical Analysis. Second Edition. John Wiley and Sons, Inc., New York, 1952, pp. 150-151.
13. Ibid.

MERIT-RATING SALARY PLANS IN PUBLIC SCHOOL SYSTEMS OF THE UNITED STATES, 1955-56

IRVIN ALBERT KARAM
Abington, Pennsylvania

ADOPTION OF merit-rating plans for teachers' salary purposes continues to be one of the most pressing and perplexing problems confronting school boards, administrators, and teachers throughout the country. The varying kinds of evidence presented in recent studies and the literature have indicated that there are many factors which condition the success of the plan, and that the addition of the merit feature to the salary schedule is, many times, more complex than the plan itself. Because of the variance of opinion, a study of merit-rating salary systems was undertaken in order to assay the basic facts underlying their development, installation, and administration.

The problem was to determine what types of merit-rating salary plans have been developed and are in operation in the public school systems in the United States. This study was limited to those public school positive merit-rating plans which were used for salary and promotional purposes to teachers already employed. This investigation does not include negative merit devices which penalize unsatisfactory service, even though negative methods are frequently found coupled with positive practices.

Purposes of the Study

1. To determine why school systems in this study adopted merit-rating salary plans.
2. To determine the nature of merit-rating salary plans.
3. To determine how these plans operate.
4. To ascertain the particular features of the plans.
5. To determine what principles seem to govern the successful operation of merit-rating salary plans.
6. To draw conclusions and make recommendations concerning what appear to be the more desirable aspects of these plans with reference to their development, features, installation, and administration.

A survey of the recent literature written by both lay and professional persons on the subject of

teachers' salaries reveals the fact that an ever-increasing amount of consideration is being given to merit rating.^{1*}

The public is increasingly demanding that the schools borrow the bonus or incentive plan from industry, tailor it to the needs of education, and thereby reward competent teachers.² This practice, they feel, would ensure a better return for the school tax dollar and would be more consistent with the principles of free enterprise for it would reward initiative and competency. Some taxpayers have indicated that they are no longer willing to support a program which automatically raises every faculty member to a high maximum salary figure.³

Further support for merit rating comes from those members within the profession itself who are of the opinion that the merit principle must be accepted by teachers generally if their salaries are to approach those of other professions and of businesses where financial remuneration is a more distinct reflection of the individual's abilities, qualities, and performance.

Other proponents contend that this system is the best method of stimulating teachers toward self-improvement, and that the teacher will approach his maximum capacity when equitably rewarded for his efforts. Also, better instruction would result because merit rating implies evaluation and supervision which at present are inadequate in many school systems. An efficient merit-rating plan would bring about closer supervision, thus revealing teachers' weaknesses and setting the stage for their correction.⁴

This school of thought also believes that competent administration can make ratings with few inequalities and that teaching can be evaluated as fairly as teachers evaluate pupils' learning.

Despite the apparent validity of these reasons, there is strong resistance from school personnel, mainly from the teachers themselves, who contend that there is no satisfactory means of measuring teaching skill. The extreme complexities of the educational process coupled with extreme variation in

* Footnotes will be found at the end of the article.

talents and competencies renders it exceedingly difficult to make valid comparisons among individual teachers. The actual contribution of an individual teacher to the overall growth and development of a child is too subjective to be reliably determined.⁵

Others who are averse to the idea feel that the morale of many teachers would be lowered, not only by lack of faith in the rating instrument itself, but by the fear of its use in the hands of autocratic and incapable administrators. Rating systems, they contend, break down the friendly relationship which should exist between teachers and administrators. They also hold that some staff members would think it necessary to curry favor with those administering the system.⁶

Further opposition to the use of merit-rating plans stems from the attitude that rating takes more time and money than the benefits derived warrant, since an efficient system would require numerous additions to the administrative and supervisory staff in order to collect and properly interpret information essential for a reliable rating.

It has also been alleged that a merit-rating plan would lead to superficial teaching so that the teacher could point to visible evidence of "merit". Conversely, there may be the danger of teachers hesitating to ask assistance from supervisors and administrators for fear that to do so would be construed as weakness.⁸

Such a diversity of opinion on the worth of merit rating in our schools seems to justify the need for an objective study, in order to examine the principles of merit rating. As a first step, the present study undertakes to discover the merit-rating salary plans in use in 1955-56 in public school systems of the United States and to describe these plans in some detail. This analysis should shed light on the numerous factors which seem to govern the operation of such plans. From this body of information, it is hoped that other students will seek further to determine the reasons for the success and failure of the various plans.

As a result of this study, useful recommendations concerning the development, installation, and administration of a merit-rating plan could be made.

Method and Procedure

A careful review of the literature on merit-rating pay plans revealed that a major portion of the writings were merely opinions either for or against the principle of merit rating for salary purposes. A surprisingly limited amount of material was available concerning the mechanics of the plans and suggested techniques for developing and administering the plans. No source was discovered which explained in detail the existing types of merit-salary plans. However, the literature disclosed the names of some public school districts which were using merit plans. With this as a beginning, a list was compiled of every public school district which was

reported to be using some form of merit rating for salary purposes.

In 1956, letters were sent to the forty-eight state superintendents or commissioners of education asking for the names of school districts within their states which were using merit-pay systems. The respondents listed the names of fourteen such districts. The salary studies of the National Education Association, the Pennsylvania State Education Association, and the literature of the United States Department of Health, Education, and Welfare contributed the major portion of the list of school districts which finally totaled 224 districts.

At this point a letter was sent to the superintendent of supervising principal of each district asking for a copy of their merit-rating salary plan, together with any supplementary information as to how and by whom the plan was developed and administered.

A total of sixty-nine or 35.2 per cent of the replies were received from districts having merit-rating salary provisions which were in active use. The superintendents of sixty-five districts replied that their salary schedules contained merit provisions which were seldom or never used. Letters from seventeen superintendents in this latter group mentioned that, although their salary schedules indicated super-maximum provisions, they had never been used because of constantly increasing maximum salaries. Twelve districts replied that the merit provision existed so that the salary schedule could be exceeded in order to prevent exceptional teachers from leaving their system. They commented further that these provisions had seldom or never been exercised.

Fifty-six or 28.5 per cent of the superintendents replied that they did not have and never did have a merit pay plan, even though it had been claimed that they did have. Six or 3 per cent of the superintendents of this group answered that they had had merit pay plans which had failed, and which were therefore discarded. Letters were immediately sent to these six superintendents asking for additional information as to the reasons for failure of the plans.

An analysis sheet was then developed listing the various items which were thought necessary to understanding and interpreting the developmental and administrative aspects of a merit-rating pay plan. Each reply from the sixty-nine school districts which were actively using their merit pay provisions was examined from the standpoint of whether or not the necessary data was provided. A second letter was sent to the superintendent of each district asking additional specific questions, the answers to which would give a more complete understanding of the plans' operation.

A third letter was sent to the sixty-nine superintendents of districts having merit-rating plans, asking for their opinions as to whether their plan was successful or not. A second question asked them to list and discuss the factors or features which they felt contributed to this success or lack of success.

A third question requested that the superintendent discuss the recommendations which they would make for a school district which is contemplating the development and installation of a merit-rating salary plan. Replies to this letter were received from fifty-six superintendents. Six superintendents refused to comment on whether their plans were operating successfully or not, but did make recommendations for developing and administering a merit-pay plan.

Summary of the Findings

The following are statements of the findings of the study:

1. Careful investigation revealed sixty-nine merit rating salary plans in active use in the United States.
2. All sixty-nine merit pay plans in the study could be divided into seven distinct types: 1) the Supermaximum plans, 2) the Accelerated Increment type, 3) the Bonus plans, 4) the Multiple-Track plan, 5) the Periodic Merit Evaluation type, 6) the Annual Outstanding Teacher Award, and 7) the Summer Merit Teacher Projects Program. On numerous occasions districts combined two or more merit features into one plan.
3. The Supermaximum type was most numerous, represented by twenty-seven districts or 39.1 per cent of the total plans in the study. The next most prevalent type was the Accelerated Increment with fifteen or 21.7 per cent of the total number. The combined Supermaximum and Accelerated Increment variety also was employed by fifteen or 21.7 per cent of the total districts. The Bonus and Periodic Evaluation type plans were each used in four or 5.7 per cent of the total. Three or 4.3 per cent of the total number were Multiple-Track plans. One district combined the Periodic Evaluation, the Multiple-Track, the Summer Merit Teacher Projects Program, and the Annual Outstanding Teacher Award into one plan.
4. The Net Current Expenditures per Pupil for districts in the study ranged from \$248 to \$742. Sixty or 87 per cent of the total plans spent amounts above the national mean for 1956, which was \$277.
5. The pupil populations in the sixty-nine districts included in the study ranged from 647 to 15,491. However, fifty-two or 75.2 per cent of the total districts had pupil enrollments of fewer than 5,000 pupils.
6. The teacher populations in the sixty-nine districts ranged from 27 to 708. Fifty-one or 73.9 per cent of the total districts had fewer than 250 teachers.
7. The mean salaries of districts using merit pay plans ranged from \$3,691 to \$6,492. Fifty-nine or 85.5 per cent of the total schedules were found to pay mean salaries above the national mean of \$4,220 for classroom teachers.
8. The salary schedule minimum ranged from a low of \$1,935 to a high of \$4,200. Fifty-five or 79.7 per cent of the total districts paid minimum salaries above the national median minimum of \$3,320.
9. Maximum salary figures in the districts studied ranged from \$4,335 to \$9,800. Fifty-nine or 85.5 per cent of the total plans were paying higher maximum salaries than the national median maximum figure of \$5,575.
10. Sixty-four or 92.7 per cent of all the plans analyzed have been developed since 1946.
11. The typical merit salary plan was found to have been developed by a committee consisting of the superintendent or supervising principal, building principal, teachers, and representatives from the school board. Teachers usually constituted a majority of the committee.
12. The majority of superintendents disclosed that merit-rating plans were developed as a result of pressure from the school board. Few merit salary plans were developed at the request of the faculty and administration.
13. The majority of merit pay plans were found to include formal written criteria for measuring teacher performance. However, most plans did not have the evaluation policies and procedures reduced to a written form.
14. The percentage of teachers receiving some form of merit pay in the sixty-nine districts in the year 1955-56 ranged from 0 to 61 per cent of the total faculty. The mean was found to be 21 per cent.
15. The most frequently reported negative merit device was the withheld increment. Mere existence of negative merit features gave no true indication of the degree of use. Few teachers were found to have been penalized when salary was determined for 1955-56.
16. The selection of teachers for merit evaluation depended largely on the type of merit plan used within the district. In the two most frequently used types, the Supermaximum and Accelerated Increments, approximately 60 per cent of the eligible teachers were rated. The remaining eligible teachers were eliminated by screening devices designed to restrict evaluation to the more likely candidates.
17. Criteria for choosing superior teachers were usually divided into several categories:
 - a. Teachers' Personal Qualities appeared as factors for determining merit in 75 per cent of all plans studied. They had a mean weighting of nine per cent of the total criteria.
 - b. Eighty-eight per cent of the plans mentioned Teaching Ability or Effectiveness as a criterion. This factor received a mean weighting of sixty per cent when used.
 - c. Pupil-Teacher Relationships were recognized as a category in 68 per cent of the plans

and had a mean weighting of eleven per cent of the total.

- d. Relationships with Staff Members appeared in 59 per cent of the rating plans and received a mean weighting of ten per cent of the total when used.
 - e. Fifty-nine per cent of the plans contained the category Contributions to the Community. The average weighting was eight per cent of the total criteria.
 - f. The factor Professional Growth was found in 39 per cent of the plans and had a mean weighting of ten per cent.
 - g. Contributions to the Total School Program appeared in 59 per cent of the plans. The average weighting was fifteen per cent of the criteria.
- Though implied in many cases, only one plan specifically mentioned Pupil Achievement as a means of appraisal. Eleven or 15 per cent of the plans had no definite rating criteria other than the subjective statement "superior teaching".
18. The typical merit-rating plan involved several persons as evaluators. Superintendents participated in the appraisal in 77 per cent of the total plans examined. In approximately 73 per cent of the systems, the building principal served as an evaluator. Department heads were participants in the appraisal process in 17 per cent of the plans. Teachers were serving as raters in only 11 per cent of the total plans.
 19. In 86 per cent of the plans, the classroom observational visit was used as a technique for determining teaching effectiveness. In only a small percentage of cases were the visits limited to just one or the other type.
 20. All the merit-pay plans provided that evidence of teacher performance be deposited in a cumulative record folder of some nature. The majority of districts employing merit-pay systems placed these files under the jurisdiction of the superintendent and building principal or the building principal alone. It was customary for the contents of these files to be confidential in nature and open to the evaluators and the teacher. In the majority of districts the teachers were encouraged to submit evidence on their own behalf.
 21. Approximately 65 per cent of the total plans contained some means whereby the teacher might appeal the decision of the evaluators. However, these privileges of appeal were seldom used.
 22. There were no public acknowledgements made of merit award winners in 71 per cent of the districts using merit pay systems. Twenty-one per cent of the districts did provide for announcement of recipients' names, but only to the professional staff. This was customarily done through the medium of superintendents or teachers association bulletins. Announcement

to the general public was made in only eight per cent of the districts.

23. Teacher opinion polls were frequently conducted by districts for the purpose of measuring teacher reaction to the merit plan. It was the usual procedure for Teachers Salary Committees, Teachers Professional Association, or ad hoc committees to meet and evaluate the effectiveness of the merit-pay plan.
24. In the majority of the districts, revision of the plan was carried out by representatives of the same groups which originally developed the plan.

Conclusions

Within the limits of this study, the following conclusions seem to be valid:

1. There are very few teachers' salary schedules in existence today which relate salary to competence.
2. The fact that the majority of the merit-salary plans judged successful by the chief administrative officials using them were in use in districts with relatively high net current expenditures, seems to indicate a positive relationship between the amount of money spent and the success of the plan.
3. The majority of the merit-salary plans judged successful by the superintendents or supervising principals using them were in use in districts paying relatively high mean salaries for teachers. This provides some evidence for the point of view that there is a positive relationship between basic salaries paid to teachers and the success of a merit plan.
4. The fact that most of the existing merit-salary plans have been originated in the past ten years probably indicates that the practice may be in the stage of initial growth.
5. Since most of the plans judged successful in the opinions of the chief administrative officials were developed by committees consisting of administrators, teachers, and school board representatives, a positive relationship is indicated between involvement of those directly concerned and the success of the plan.
6. Most merit-pay plans seem to have been developed as a result of the wishes of the school board.

Recommendations for Developing a Merit-Pay Program

The following question was asked of the chief administrative officials in districts presently employing merit-pay plans: "What recommendations would you make for a school district which is contemplating the development and installation of a merit-rating salary system?" The following is a summary of the most frequent replies:

1. Those persons who will be most directly affect-

- ed by a merit-rating plan should share in its development. In formulating a plan, teachers should be well represented.
2. The faculty and administration must recognize a need for a properly administered merit-pay plan and be convinced of its possible benefits.
 3. For a plan to be successful there must be mutual confidence and respect between the teachers and those who administer the plan.
 4. Actual plan construction should be preceded by ample research on the subject. Plans that are presently in operation should be studied in light of local needs, but not borrowed in their entirety. The plan should definitely be locally developed.
 5. The community, through its representatives on the school board, must establish a range of teachers' salaries that reflects the importance of the job. The basic salary schedule must be adequate with the average teacher being able to earn at least as much as the average teacher in a district not using a merit-pay plan.
 6. The merit-pay plan should include a philosophy or statement of purposes. This philosophy should reflect the thinking of the majority of those concerned.
 7. There should be an almost universal agreement within the district among school personnel on the criteria for measuring teacher performance.
 8. Ample time should be set aside for the appraisal of teacher performance.
 9. Attention should be given to improving physical facilities and working conditions. The teacher must be provided with a teaching situation that is conducive to learning.
 10. The merit increase should be sufficient in amount to provide an incentive.
 11. The major purpose of any merit-rating plan should be to improve the caliber of teaching. If this is to be done, then supervision must be continuous and extensive. The results of teacher appraisal should be used to improve teaching.
 12. There should be a positive approach on the part of those persons who administer a merit-rating plan. The emphasis should be on improving instruction rather than penalizing unsatisfactory teachers.
 13. Some form of rating sheet should be used during the evaluation to channel the rater's thinking.
 14. A merit-rating salary system should not be designed to produce uniformity of teaching

methods.

15. To increase effectiveness, the plan must provide an incentive for teachers at all levels of experience.
16. Merit-pay plans should be continually evaluated and their effectiveness appraised.
17. A merit-rating salary plan should include a provision for revising the plan. Representatives of all parties directly affected by the plan should participate in the revision.

Recommendations for Further Research

This research has been a status study. Further investigation in the field of merit systems would seem almost unlimited; however, recommendations proposed here are those related to the present study.

1. It would be valuable to employ depth interview techniques with teachers, administrators, school board members, and other personnel directly involved in the merit system so that more extensive information concerning the underlying factors responsible for success and failure of merit-pay plans could be collected.
2. The field of industrial merit rating should be explored from the standpoint of its implications for educational merit-pay plans.

FOOTNOTES

1. Staub, Frederick, and Savage, William W., "Teachers Salaries and Merit Rating," Administrators Notebook, Vol. IV, University of Chicago Press, May, 1956, p. 1.
2. Rogers, Virgil M., "Appraising Teacher Efficiency for the Betterment of Schools," The School Executive, Vol. 67, April, 1948, p. 54.
3. Report of the Professional Personnel Policies Committee of the Board of Education of Manhattan, New York, December, 1955, p. 1.
4. Department of Research, San Diego City Schools San Diego, California. Merit Salary Programs for Teachers, September 15, 1953, p. 2.
5. Department of Classroom Teachers of the National Education Association. Summary Statement of Study Conference on Merit Rating, Washington, D.C., November 23-24, 1956.
6. Greenwald, Catherine I., "One City's Disaster in Merit Rating," American Teacher, Vol. 22, May, 1956, p. 12.
7. Department of Classroom Teachers of the National Education Association, op. cit., p. 3.
8. Ibid., p. 4.

AN INVESTIGATION OF THE EFFECTS OF A SEVENTH AND EIGHTH GRADE CORE PROGRAM

BERNARD SCHWARTZ
Pennsbury Schools, Fallsington, Pennsylvania

WITHIN THE PAST two decades, many new instructional ideas have appeared upon the educational scene and, though these ideas have differed in their approach to curriculum and method, they have all revolved around one aim, that of fulfilling more effectively the educational needs of children. One of the most promising of the newer approaches to curriculum has been labeled the "core program" approach.

Recognizing the crucial need for obtaining factual evidence about the effects of the core program in the educational field, an attempt was made in this study to shed some light on the relationship between participation in a core program and success as a high school student. The initiation and operation of a core program thrusts heavy burdens upon the shoulders of school administrators and teachers. The transformation from the traditional teaching system to the core system greatly affects all who are either directly or indirectly involved in the field of education. Therefore it is cogent to evaluate the effects that the core program produces upon its students, as compared with those elicited by the conventional method.

The Problem

The major purpose of this study was to evaluate the effectiveness of the core program at Pennsbury High School, Yardley, Pennsylvania. Specifically, it was hypothesized that the following characteristics would be noted when children who have learned through a core program in junior high school were compared in senior high school with an equated group of students who did not learn through the core program:

1. High school students who have learned through a core program will achieve significantly higher scores in various selected subject matter and skill areas as measured by standardized achievement tests.
2. High school students who have learned through a core program will achieve higher final marks in grades ten, eleven and twelve.

3. High school students who have learned through a core program will exhibit more favorable personality and behavior traits when rated by their teachers.
4. High school students who have learned through a core program will participate in a greater number of extra-curricular activities, a greater variety of extra-curricular activity, and will devote more time to these activities.
5. High school students who have learned through a core program will maintain or increase their advantage over non-core experienced students in all of the above factors, as they progress from grade ten to grade eleven, and then to grade twelve.
6. High school students who have learned through a core program will exhibit more favorable social attitudes and more favorable attitudes toward various aspects of school life as measured by a standardized student attitude scale.

The Population

The population for this study consisted of the 236 seniors in the Pennsbury High School graduating class of 1957. Of this number, 168 met the requirements established for inclusion in the study. These requirements were: 1) availability of either sixth grade class marks, or of some acceptable standard indication of intelligence level in the sixth grade; 2) attendance at Pennsbury High School for grades ten to twelve inclusive; and 3) availability of data on the various factors investigated. Of the 68 students excluded from the study, 45 were eliminated because it was not possible to secure either their sixth grade school marks or an intelligence quotient obtained in the sixth grade. Also excluded were 19 students who entered Pennsbury High School after the beginning of grade ten. Four students were eliminated after being selected for the study because their test results on important factors, selected for comparison of the two groups, were missing.

The 168 students ultimately used in the study were then divided into two groups. The first group

had all participated in the core program in grades seven and eight at Pennsbury. This experimental group consisted of 80 students, 38 males and 42 females. The students of the second group, which shall be referred to as the control group, did not have core program experience. This control group totaled 88 students, 42 males and 46 females. For the purpose of this study it was assumed that they received instruction of equal quality in grades seven and eight although they came from different school systems in Pennsylvania and other states.

Equating the Control and Experimental Groups

The entire core group had been in the Pennsbury Schools from grades seven to twelve. Most of the control, or non-core group entered Pennsbury in grade nine. A small number of non-core students entered Pennsbury at the beginning of the tenth grade. It was deemed advisable to eliminate ninth grade data in making comparisons between the control and experimental groups. It was assumed that the results of such comparisons would be spurious because the core group was better acquainted with the high school, since it is actually an extension of the junior high school, in school climate, tone, philosophy, and even in physical plant. To minimize the core group's advantage in school orientation, group comparisons were not made for the ninth grade. This gave the non-core students a year to become adjusted to the school and its manner of operating.

It is important to note that although the experimental group had participated in the core program in grades seven and eight, they were intermingled with non-core students in the high school classes. Thus the homeroom and instructional classes of the tenth, eleventh, and twelfth grades consisted of both experimental and control group students. In this manner those designated as non-core students had the same teachers and received the same instruction and attention as did the core students in the same class. This eliminated the possibility of attributing differences found between the two groups in high school to special instruction or attention on the part of the teachers.

Briefly, the two groups were equated on the following factors:

1. Number within the group.
 2. Sex.
 3. Chronological age.
 4. High school instruction.
 5. High school areas of concentration.
 6. Intelligence (measured in grade eleven.)
 7. Sixth grade evaluations.
 - a. School marks.
 - b. Standardized achievement test results.
 - c. Intelligence quotients.
 8. Socio-economic status.
 - a. Educational level of parents.
 - b. Occupation of chief wage earner.
- No statistically significant differences were

found between the control groups on any of the above factors. Substantial similarity between the core and non-core groups was established for each of these factors.

Factors Investigated

It was the aim of the investigator to make statistical comparisons between the control and experimental groups on as many relevant factors as possible. The factors selected for investigation and their sources of data were as follows:

1. Subject matter and skill areas achievement in high school as measured by standardized tests. The core and non-core groups were compared on each of the nine sub-tests comprising the Iowa Tests of Educational Development. This battery was administered twice to all the participating students in the tenth grade and again in grade twelve. Results of the California Reading Test, administered in grade eleven, were also analyzed. Comparisons between the two groups were made for scores obtained in the reading vocabulary section, the reading comprehension section and the total reading score.
2. High school report card marks. Final report card marks, listed numerically, for grades ten, eleven and twelve were obtained from the school files. The control and experimental groups were then compared for each year by averaging the marks achieved in the five major subjects carried by each student. In addition, separate comparisons were made of the two groups on marks achieved each year on the two subjects which all students were required to take--English and social studies.
3. Personality and behavior traits. In grade ten and again in grade twelve all participating students were rated on a variety of personality and behavior traits. The ratings were performed for each student by the five teachers he had in his five major subjects. By assigning numerical weights to each step within each attribute scale, it was possible to adapt these evaluations for statistical analysis. Comparisons were then made between the control and experimental groups for each of the six personality and behavior characteristics rated in the tenth grade and for each of the seven attributes rated in the twelfth grade.
4. Amount and extent of extra-curricular participation in high school. Three approaches were employed in determining differences between the core and non-core groups on the topic of extra-curricular participation. The first approach involved noting the number of activities in which each student participated each year. The second appraisal involved examination of the types of extra-curricular activities in which the students were participants. The final approach entailed determining the amount of time (hours per week) the students devoted toward extra-curricular

participation. Employing the last approach it was possible to compare the control and experimental groups in grade twelve only. However, group comparisons in grades ten, eleven and twelve were made for the first two types of extra-curricular activity appraisals.

5. Determination of the degree to which the investigated factors are maintained through each high school grade. It was deemed important to ascertain whether results and significant differences noted in the tenth grade comparisons of core and non-core groups also persisted through grades eleven and twelve. Included in this phase of the investigation were: results of standardized achievement tests in subject matter and skill areas; high school marks; four of the rated personality and behavior traits, and number and types of extracurricular activities participated in by the students.
6. Social and school attitudes. The investigator adopted and revised a questionnaire for the purpose of obtaining indications of the social adjustment and school attitudes of the participating pupils. The following four areas were derived by further adaptation of the attitude scale: 1) personal-social adjustment, 2) attitude toward classmates, 3) attitude toward teachers, and 4) attitude toward school administrators and school operation. An appropriate scoring system was devised by the investigator. Comparisons between the control and experimental groups were made on the total scale scores and on the scores for each of the four established areas.

Findings

In the attempt to obtain valid information about the established hypotheses, appropriate statistical procedures were applied to the data. The results are presented in the light of statistically significant differences obtained by comparing the experimental and control groups on the various factors investigated.

1. Subject matter and skill areas. The results of the standardized achievement tests used in the study revealed that:
 - a. The results of the Iowa Tests of Educational Development, administered in grade ten, revealed that the core group achieved significantly higher scores in Reading -- Natural Science and General Background -- Natural Sciences.
 - b. The results of the same battery of tests, administered again in grade twelve, showed that the experimental group's scores were significantly higher in Quantitative Thinking, as well as Reading -- Natural Science and General Background -- Natural Sciences.
 - c. Statistically significant differences in favor of the experimental group were also found for the Total Vocabulary and Total Reading scores of the California Reading Test admin-

istered in grade eleven.

2. High School Report Card Marks. An examination of report card marks revealed:
 - a. The core group received significantly higher final averages in grades ten, eleven and twelve.
 - b. The core group received significantly higher final marks in the required social studies courses in grades ten, eleven and twelve.
 - c. The core group obtained significantly higher final marks in the required English courses in grades ten and twelve. No statistically significant difference was found in grade eleven in this area.
3. Personality and Behavior Traits. Personality and behavior trait findings revealed that:
 - a. The core group, both in grades ten and twelve, received significantly higher ratings in Responsibility, Influence, and Seriousness of Purpose.
 - b. The experimental group members rated significantly higher in grade ten in Creativeness and Adjustability.
 - c. Statistically significant differences in favor of the core group were found in grade twelve in Industry and Initiative.
 - d. The two groups received similar ratings in both grades in the attributes Concern for Others and Emotional Stability.
4. Extra-Curricular Participation. The topic of extra-curricular participation was analyzed by employing three distinct approaches in making experimental and control group comparisons. The following results were obtained by these approaches:
 - a. Number of activities participated in -- The core group engaged in a significantly greater number of extra-curricular activities in grades ten and eleven. No statistically significant difference was found in grade twelve.
 - b. The two groups did not differ in the types of extra-curricular participation (e.g. athletic, honorary, social, etc.) in any of the three investigated grade levels.
 - c. The twelfth grade investigation the number of hours per week devoted to extra-curricular activities by experimental and control students revealed no statistically significant differences.
5. Degree of Maintenance of Factors Through High School Grades. Results of this phase of the study indicated that:
 - a. Statistically significant superiority exhibited by the core group in Natural Science in grade ten was maintained in grade twelve.
 - b. The significantly higher total final averages achieved by the core group in grade ten persisted through grades eleven and twelve.
 - c. The core group's statistical superiority in social studies final marks in grade ten was maintained through grades eleven and twelve.
 - d. The core group's superiority in English

marks in grade ten, though not maintained through grade eleven, did reappear in grade twelve.

- e. Of the four personality and behavior traits which were measured, the core group achieved significant superiority in both the tenth and twelfth grade ratings of Seriousness of Purpose, Influence, and Responsibility. The other attribute, Concern for Others, was not found to be significantly different between the two groups in either the tenth or twelfth grade ratings.
6. Social and School Attitudes. No statistically significant differences appeared between the experimental and control groups when total scores of the social and school attitude scale were examined. This was also the case when the two groups were compared on scores of each of the four sub-sections of the scale.

Conclusions

Within the limitations of this study, the following conclusions, based upon statistical findings, can be drawn regarding the specific hypotheses established:

1. The hypothesis that high school students who have learned through a core program in junior high school achieve significantly higher scores in various subject matter and skill areas as measured by standardized tests can be accepted in part.
2. The hypothesis that high school students who have learned through a core program in junior high school achieve higher final marks in grades ten, e-

leven and twelve can be accepted.

3. The hypothesis that high school students who have learned through a core program in junior high school exhibit more favorable personality and behavior traits when rated by their teachers can be accepted in part.

4. The hypothesis that high school students who have learned through a core program in junior high school participate in a greater number of extra-curricular activities can be accepted. However, the hypotheses that they participated in a greater variety of extra-curricular activity and that they devote more time to these activities must be rejected.

5. The hypothesis that high school students who have learned through a core program in junior high school maintain or increase their advantage over non-core experienced students in the above factors as they progress from grade ten, to grade eleven, and then to grade twelve can be accepted in part.

6. The hypothesis that high school students who have learned through a core program in junior high school exhibit more favorable social attitudes and more favorable attitudes toward various aspects of school life as measured by a standardized student attitude scale must be rejected.

FOOTNOTE

The material presented in this paper is based on the author's doctoral dissertation accepted in August, 1958 by Temple University. A more complete analysis of the data may be found upon examination of the dissertation.

A STUDY OF STUDENT DISCIPLINARIAN PRACTICES IN TWO GEORGIA HIGH SCHOOLS

KARL C. GARRISON
University of Georgia, Athens

ALTHOUGH DISCIPLINE is an old and universal problem, one still finds conflicting claims and viewpoints about it. The word "discipline" is derived from a Latin verb signifying "to learn." Thus, the values claimed for discipline relate closely to learning.

The patterns of behavior of children are not always in accord with those deemed desirable by adults. Generally speaking, children are steered into certain preferred ways of behaving through a system of rewards and punishment. One of the uses of discipline is to teach the child a reasonable degree of social conformity. Through discipline the child is taught to wear clothes, not to expose certain parts of the body, to keep clean, and to eat certain foods in a prescribed manner. As he reaches high school, he is taught that he must conform to certain rules and practices of the school, he must give attention to his studies, and must inhibit certain impulses in order that a favorable learning situation may exist. Certainly one of the major problems facing most high school teachers is that of dealing with misbehavior and motivating students to behave in certain desired ways.

In order better to evaluate discipline practices and the reaction of students to such practices, it was considered desirable to secure first an evaluation of their delinquency proneness. The KD Proneness Scale¹ was administered to all students of the two high schools studied. The results, presented in Table I, show the mean scores of students from the high schools were slightly lower than those from whom norms were developed in obtaining standard scores for the Scale. A comparison of mean scores from the two schools revealed no significant difference. Thus, the data from the two rural and small town schools were treated as combined data. The group of students studied may be regarded as a group slightly below average, based upon norms for the Scale, in delinquency proneness.

Complete data were available on 276 high school students distributed by grades as follows: ninth grade, 86; tenth grade, 70; eleventh grade, 63; and

twelfth grade, 57. A schedule adapted from the Illinois Inventory of Public Opinion was developed and administered to the 276 students of this study. The purpose of the schedule was to secure factual information about punishment from each student along with opinions about various school situations, with particular reference to matters relating to discipline.

Information reported by the students relative to recent behavior for which they were reprimanded or punished is presented in Table II. The seven most frequently reported misbehaviors, stated in order of decreasing frequency for the total group, are: talking, 45.6 per cent; inferior school work, 5.1 per cent; disturbing class, 4.4 per cent; inattention, 4.1 per cent; laughing, 4.1 per cent; tardiness, 3.7 per cent; and cutting class, 3.7 per cent. Obviously, talking is the most frequent source of teacher's displeasure, accounting for more than 45 per cent of the behaviors for which students were corrected.

A comparison of the self-reported misbehaviors of girls and boys revealed several significant sex differences. Approximately 60 per cent of the girls were corrected for talking as contrasted with approximately 38 per cent of the boys. Disturbing the class and inattention were misbehaviors more frequently reported by the boys, while the girls reported laughing and tardiness more frequently.

The types of punishment most frequently administered, in order of decreasing frequency for the total group, are: reprimanded before the group, 42.8 per cent; reprimanded in private, 15.6 per cent; kept after school or during recess, 11.2 per cent; given special assigned task, 9.2 per cent; sent from room, 5.8 per cent; assigned special seat, 4.1 per cent; sent to principal, 2.8 per cent; grade lowered, 2.0 per cent; and miscellaneous, 5.7 per cent.

Another feature of this study was to obtain an appraisal from the students of the administration of discipline. Of extreme significance are the facts that 72.8 per cent of the girls and 78.6 per cent of the boys readily admitted guilt for the offenses for which they were punished. Over one half (52.4 per

*Footnotes will be found at the end of the article

TABLE I
COMPARISON OF THE SCORES ON THE K D PRONENESS SCALE OF STUDENTS
FROM TWO GEORGIA HIGH SCHOOLS WITH THOSE OF THE NORM GROUPS

Group	Mean	S. D.
Georgia High School Boys	-14.10	6.24
Georgia High School Girls	-12.60	6.03
High School Norms (Grades 10-12) Boys	-12.22	7.40
High School Norms (Grades 10-12) Girls	-11.77	6.39
"High Morale" Group-Boys	-14.52	6.87
"High Morale" Group-Girls	-13.51	5.87
"Delinquent" Group-Boys	+ 2.75	8.29
"Delinquent" Group-Girls	+ 5.67	7.54

A plus score is indicative of delinquency proneness, while a high minus score is indicative of low delinquency proneness.

TABLE II
MOST RECENT BEHAVIOR FOR WHICH STUDENTS WERE
CORRECTED, AS REPORTED BY STUDENTS

Behaviors	Girls (141)	Boys (135)	Total (276)
Talking	84	51	135
Inferior school work	6	9	15
Disturbing class	3	10	13
Inattention	2	10	12
Laughing	9	3	12
Tardiness	7	4	11
Cutting Class	3	8	11
Violating school rules	3	7	10
Impudence	4	6	10
Passing notes	5	4	9
Throwing objects	1	8	9
Carelessness	3	6	9
Chewing gum	6	2	8
Disobedience	2	5	7
Fighting		5	5
Daydreaming	4	1	5
Running in hall		4	4
Using obscene language		3	3
Signing mother's name to paper	2	1	3
Improperly dressed	2		2
Stubbornness	1		1
Pinching others		1	1

cent) of the girls and 63 per cent of the boys stated that the punishment was justifiable, while 17.1 per cent of the girls and 15.2 per cent of the boys felt that they had been punished unfairly. A small percentage of the students (girls, 5.8 per cent; boys, 9.3 per cent) indicated that the teachers were too strict. A smaller percentage (girls 4.2 per cent; boys, 1.4 per cent) felt that discipline was too lax. Approximately three fifths of the girls and two thirds of the boys were of the opinion that the teachers were not consistent. This was indicated by such remarks as:

"A teacher gets after one student for 'acting up' but lets another student get by."

"They show partiality to certain students, mostly depending upon who their parents are."

"They 'ride' you when they are in a bad mood."

"They show partiality to the most popular students."

"They sometimes accuse the wrong person, and won't believe you."

The results of this study point up the frequency of students misbehavior as a source of annoyance to the teacher. Talking and laughing on the part of girls was frequently reported. In addition to talking, boys were more frequently given to more aggressive forms of misbehavior such as fighting, throwing things, and the like. Reprimanding the students was the most frequent type of punishment used by the teachers. Other forms of punishment frequently used were keeping students in after school or during the recess period, giving them

special assigned tasks, assigning them another seat, having them report to the principal, or lowering their grades.

Although the students were in general agreed that punishment was justifiable, the majority of them felt that the teachers were not consistent in their discipline. This is a serious indictment in light of the importance attached to consistency in discipline in relation to character development. Results from an eight-year study of the development of moral character in a group of children revealed a correlation coefficient of .62 between consistency of discipline and moral competence.² There is much evidence to show that consistency of discipline is more important than either severity or laxity. The concern of the students of this study was over consistency with most of the students indicating that the punishment was neither too strict nor too lax.

FOOTNOTES

1. Kvaraceus, William C., K D Proneness Scale. World Book Company, Yonkers-on-Hudson, 1950. The items in the K D Proneness Scale were derived from those areas in which significant differences between delinquents and non-delinquents have been reported in the literature. Evidences are presented indicating the reliability and validity of the scale.
2. Havighurst, Robert J., "The Function of Successful Discipline," Understanding the Child, 1952, Vol. 21, pp. 35-38.

SOME RESULTS OF AN ENRICHMENT PROGRAM FOR GIFTED NINTH GRADERS

J. A. R. WILSON
University of California, Santa Barbara College

Historical Orientation

In 1952-53 in one of California's junior high schools a committee was set up to discuss a program for more capable students. As a result of these discussions it was decided to select a group of sixteen students from the top ninth grade classes, teach them algebra approximately three days a week and use the remaining time for an enrichment program. This program was to consist of the preparation and presentation of individual research reports, visits to and note-taking at college lecture series, field trips to local industries, and the possibility of the informal discussion that is feasible in a class of sixteen.

The ninth grade class of 1953-54 was the first to operate under the proposed outline. Check was made of the success of this class in algebra and of the teacher's and students' subjective reactions. These appraisals were all highly favorable. Classes have been organized on the same basis each year since.

As the first group finished high school, it seemed desirable that an objective statistical study of results should be undertaken. This paper is the result of that study.

Purposes

1. To attempt to see whether a shortened period of time spent on ninth grade algebra was harmful to further work in mathematics.
2. To attempt to see whether the "seminar" periods which replaced the algebra periods produced measurable improvement in:
 - a. Academic school work other than mathematics.
 - b. Leadership roles as indicated by membership and office holding in clubs and activities both in and out of school.

Design of the Study

1. Selected the special class that had been in the ninth grade in 1954-55 and who, at the time of the study, were high school seniors.

2. Developed a comparison group of other students who had been classmates in 1954-55.

3. Statistically compared the experimental group with the comparison group by using the significance of the differences of the means for small correlated groups, as set forth by Guilford, on the following points:

- a. Intelligence test scores.
- b. Eighth grade achievement scores in reading.
- c. Eighth grade achievement scores in arithmetic.
- d. Seventh and eighth grade academic grades.
- e. The Lee algebraic test.

4. Administered the Cooperative Algebra Test (level 1) to the seniors.

5. Calculated the significance of the difference of the means for the results of the Cooperative Algebra Test.

6. Calculated the significance of the differences of the changes of:

- a. 7th and 8th grade mathematics scores and 10th grade geometry scores.
- b. 8th grade arithmetic achievement test scores and 10th grade arithmetic test scores.
- c. 7th and 8th grade academic grades and 10th 11th and 12th grade academic grades (excluding mathematics).
- d. 8th grade reading scores with 10th grade reading scores.

7. The election of courses by subject matter areas in 10th, 11th and 12th grades as follows:

- a. Mathematics
- b. Science
- c. Social Studies
- d. Foreign Language
- e. English
- f. Fine Arts
- g. Practical Arts

8. Leadership roles evaluated on the basis of the following tabulation: (Count each year separately)

Membership in a club	1
Executive of a club	add 1
President of organization	add 2
State or national executive	add 3
State or national officer	add 4

The differences in the groups in leadership were calculated as they were seen in the 10th, 11th and 12th grades. The changes were also calculated from the 7th and 8th to the 10th, 11th and 12th grades.

9. The members of the special class were asked to express themselves as they saw the results of the class four years later.

10. The material was reworked, omitting the scores of a gifted non-achiever and her comparison.

Problems in the Study

1. The small size of the groups (13 in each section - 3 of the original have moved away).

2. The inclusion in the experimental group of a gifted non-achiever who remained a non-achiever and who was paired with a less gifted non-achiever who became much more of an achiever.

3. The unique individuality of highly capable students.

Observations

1. The comparison group was selected by the counselors at the school on the same basis as was the original experimental group. The basis of selection was:

- IQ with a target minimum of 130.
- 8th grade reading achievement scores on the CAT.
- 8th grade arithmetic achievement scores on the CAT.
- Scores on the Lee Algebra Test administered in the 8th grade.
- Academic achievement in the 7th and 8th grades.

The comparison group was selected on a man for man basis. It is impossible to select a comparison group that exactly matches a previously chosen group selected as "the most capable learners." It was possible to get a reasonable approximation. This fact may be realized from an examination of the statistics of comparison of the two groups which follows.

2. Statistics of comparison between experimental and comparison groups. --Differences of means are obtained by subtracting scores obtained by the comparison group from those of the experimental group. A positive score is favorable to the experimental group and a negative score is favorable to the comparison group.

- Intelligence Test scores (judged most realistic from approximately 5 scores on CMM).

Group	Range	Mean	Standard Deviation
Experimental	122-150	132	7.3
Comparison	118-144	128	7.4

The difference of these means is 4 IQ score points. The standard error of the difference is 1.6 which gives a t of 2.5 significant above the 5% but below the 1% level of confidence. Reworking this material

for 12 pairs of students reduces t to 2.35 still above the 5% level of confidence. Other scores remain substantially the same.

- Reading scores on the California Achievement Test.

Group	Range	Mean	Standard Deviation
Experimental	9.4-11.5	10.55	.5
Comparison	9.6-10.8	10.3	.4

The difference of these means is .25. The standard error of the difference is .12 which gives a t of 2.1 or almost significant at the 5% level of confidence. Reworking this material for 12 pairs leaves the range mean and standard deviation the same but reduces t to 1.9 which is below the 5% level of confidence.

- Arithmetic scores on the California Achievement Test.

Group	Range	Mean	Standard Deviation
Experimental	10.1-12.0	10.7	.5
Comparison	9.5-10.9	10.1	.4

The difference of these means is .6. The standard error of the difference is .14 which gives a t of 4.3 or significant well above the 1% level of confidence. Reworking this material for 12 pairs leaves the range, mean and standard deviation substantially the same but reduces the t to 4.0 which is still significant well above the 1% level of confidence.

- 7th and 8th grade academic grades based on A-4, B-3, C-2, D-1, F-0 for academic subjects on the average of two semester grades each. Possible 8.0.

Group	Range	Mean	Stand. Dev.
	13 12	13 12	13 12
Experimental	5.6-8.0	7.1-8.0	7.5-7.6 .67 .3
Comparison	5.1-7.9	6.1-7.9	6.9-7.1 .78 .57

- Lee Algebra Test scores.

Group	Range	Mean	Stand. Dev.
Experimental	104.1-144.6	127.5	11.0
Comparison	86.1-126.8	119.	11.0

The difference of these means is 8.5. The standard error of the difference is 4.5 which yields a t of 1.9 not quite significant at the 5% level of confidence. Reworking the material for 12 pairs yields the same range and standard deviation, the mean becomes 118.7 for the comparison and 127.5 for the experimental group. The difference is 4.9 which yields a t of 1.8 which is not significant at the 5% level of confidence. As was to be expected, a study of the experimental and the control groups indicates that the experimental group is better than the comparison group. Projecting these differences, it would seem that the experimental group should improve slightly more than the comparison group during the ensuing years.

- Analysis of results of administering the Cooperative Algebra Test.

Group	Range	Mean	Stand. Dev.
	13 12	13 12	13 12
Experimental	48-80	50-80	67 68.5 10.3 9.2
Comparison	66.79	66 79	73 73.3 4.7 4.8

The difference of these means is -6. The standard error of the difference of the means is 2.9 or a t of

-2.1, almost significant at the 5% level of confidence. Note: This difference is favorable to the comparison group. Reworking this material for 12 pairs of students gives a difference of -4.8, a standard error of the difference of 2.8 yielding a t of -1.7. Not significant at the 5% level of confidence.

4. Calculations of the significance of the differences of changes from:

a. 7th and 8th grades to 10th grade geometry. The experimental group went down .15 score points and the comparison group went up .25 score points. This is a net difference in means of -.40. The standard error of the difference of the means is .66 which gives a t of -.6 which is not significant. Note it is a negative difference. Reworking this material with 12 pairs of students shows that the experimental group increased the average mean by .12 score points while the comparison group increased .08 score points giving a net difference of .04. The standard error of this difference is .6 which gives a t of .06 which is not significant.

b. 8th grade arithmetic achievement scores compared to 10th grade arithmetic achievement scores. The experimental group increased its scores 4.3 g.p. scores and the comparison group increased its scores 4.7 score points. The difference is -.4. The standard error of the difference of these means was .29 yielding a t of -1.6. This is not a significant difference. Reworking this material for 12 pairs of students the experimental group increased its score 4.3. The comparison group increased its score 4.6. The difference of these increases is -.3. The standard error of the difference of these means is .27 which yields a t of -1.1 which is not significant.

c. 8th grade reading compared to 10th grade reading. The experimental group increased its score on the average 3.15 score points and the comparison group increased 2.95 points. The standard error of the difference of the means was .27 yielding a t of .7 which is not significant. Reworking this material for 12 pairs of students the experimental group increased its mean score 3.1 g.p. scores, the comparison group increased 3.05 g.p. scores for a mean difference of .05, a standard error of the difference of .26 yielding a t of .02 which is not significant.

d. Academic grades other than mathematics were compared as they changed from 7th and 8th grades to 10th, 11th and 12th grades. Basis used A-4, B-3, C-2, D-1, F-0 with scores adjusted to two semester average grade, making a possible maximum of 8. The mean difference of the scores was .25. The standard error of the difference was 3.2 giving a t of .08. When these scores were reworked on the basis of 12 pairs, the mean difference of the score was .5, the standard error of the difference was 2.0 giving a t of .25. Neither of these t scores is significant.

5. The election of courses by subject matter areas in the 10th, 11th and 12th grades.

a. Mathematics: The experimental group elected 7 year courses or an average of .54 fewer years of mathematics than did the comparison group. The standard error of the difference of these means is .26. This yields a t of -2.1 which is almost significant at the 5% level. When these choices are reworked for 12 pairs, the average difference becomes -.5 years. The standard error of the difference of these means is .3 which yields a t of -1.7 well below the 5% level of confidence.

b. Science: The experimental group offered one less science course or an average .1 year less science than the comparison group. The standard error of the difference of the means was .28 which yields a t of -.4 not significant. When these choices are reworked for 12 pairs, there is no substantial change.

c. English: The experimental group and the comparison group offered on the average the same number of English courses. This statement is also true when reworked for 12 pairs of students.

d. Social Studies: The experimental group offered two courses or an average of .2 years less social studies than the comparison group. The standard error of the difference of these means is .26 giving a t of .08 which is not significant. Dropping one pair from the comparison does not substantially change these figures.

e. Foreign Language: The experimental group offered 3 years less or an average of .23 years less foreign language than the comparison group. The standard error of these differences is .26 yielding a t of -.8 which is not significant. Reworking the material on the basis of 12 pairs eliminates one of the years of difference and reduces the t to -.7.

f. Fine Arts: The experimental group offered on the average, .69 years more fine arts than the comparison group. The standard error of the difference of these means was .47 which gives a t of 1.5. This is not significant at the 5% level of confidence. Reworking this material on the basis of 12 pairs of students does not affect these figures materially.

g. Practical Arts: The experimental group offered, on the average, .46 years more practical arts than did the comparison group. The standard error of the difference of the means is .34 which yields a t of 1.35, not significant at the 5% level of confidence. Reworking figures on the basis of 12 pairs of students does not affect this result materially.

A summary of this material would be that the experimental group apparently took significantly fewer courses in mathematics, insignificantly fewer courses in science, foreign language and social studies and somewhat more courses in fine and applied arts.

6. Leadership roles as evaluated for the 10th, 11th and twelfth grades. The experimental group received 256 points for leadership as against 319 for the comparison group. On the average the experimental group received 4.8 fewer points. The

standard error of the differences of these means was 4.3 which yields a t of -1.1 . This is not a significant difference.

7. When the changes in the differences in leadership from 7th and 8th grades to 10th, 11th and 12th grades are studied, the mean of the differences is -1.6 score points. The standard error of the difference is 4.3 which yields a t of $-.4$. This is not significant. When these figures are reworked for 12 pairs the mean of the difference becomes -2.3 . The standard error of the difference becomes 4.6 yielding a t of $-.5$ which is not significant.

8. Subjective evaluation by the special class: With one exception (a boy whose grades have not held up in high school) the special class replied that they felt the experience has been very good for them. They felt that taking algebra 3 days a week rather than 5 had been a more stimulating experience than it would have been had they been in a class with slower students where it was necessary to repeat explanations for those students who had not understood the material the first time. They felt that the seminar experience had helped them in other subjects in high school and as individuals. When students of both groups were asked why they had not taken more mathematics, those who did not continue mathematics beyond plane geometry replied that they had not had time to fit it into their programs but that they were planning to take more mathematics in college. Two girls replied that they had not enjoyed geometry and had not considered mathematics necessary for their future occupations (elementary teacher and nurse). One, a gifted non-achiever, wrote "I stopped with geometry because I despise math. I couldn't see where it would help me or why I needed it." Almost all of the students took the time to mention the teacher by name and to point out that in their estimation he had been responsible for the course being such an outstanding one as far as they were concerned.

Conclusions

Re Mathematics:

1. The experimental group did not do as well on the cooperative algebra test as would have been expected. This statement is still probably true even when the non-achieving student is eliminated.

2. The experimental group took significantly fewer mathematics courses. For the twelve achievers there is some doubt about the significance of the smaller number.

3. The change from 7th and 8th grade mathematics to 10th grade geometry favored the comparison group although not at a significant level of confidence. There is no difference in the groups when 12 pairs of students are considered.

4. The results of the achievement tests at the 8th and 10th grade levels favored the comparison group although not at a significant level of confidence. There was even less difference when 12 pairs were studied.

Re Academic School Work:

There was no measurable difference between the two groups in academic work other than mathematics. The figures slightly favor the experimental group.

Re Leadership:

There is no measurable difference in leadership as this was defined and measured.

Summary

The statistical treatment of the results of this program for this class seem to be less favorable than the subjective evaluation had indicated would be the result. It must be emphasized that the doubt raised by this study of the desirability of this approach to enrichment only raises a question. A judgement must wait for other studies that will study other groups operated in a similar context.

JOURNAL OF EDUCATIONAL RESEARCH

JANUARY 1960

Volume 53

Number 5

CONTENTS

Fifty Seventh Graders: A Comparison of Their Reading Achievement and Expected Achievement in Grades One Through Seven— Jeanne S. Hoyt and Dorothy S. Blackmore.....	163
A Comparative Study of Achieving and Underachieving High School Boys of High Intellectual Ability—Edward Frankel.....	172
An Empirical Study of Performance in Mathematics and Performance on Selected Entrance Examinations—Joseph R. Marches....	181
An Investigation of an Experimental First Grade Phonics Program—Richard H. Bloomer.....	188
The Relationship of Colors to Various Concepts—Harry S. Beck....	194
Research Insights into the Relationship Between Teachers' Acceptance Attitudes, Their Role Concepts, and Students' Acceptance Attitudes—Robert E. Cummins.....	197
Changes in Ethnic Reaction Tendencies During High School—John D. McNeil.....	199

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

FIFTY SEVENTH GRADERS: A COMPARISON OF THEIR READING ACHIEVEMENT AND EX- PECTED ACHIEVEMENT IN GRADES ONE THROUGH SEVEN

JEANNE S. HOYT and DOROTHY S. BLACKMORE
Dominican College of San Rafael, California

Purpose of the Study

Reading achievement of children should normally be expected to progress grade by grade in a somewhat systematic order in terms of intelligence and interest, depending upon such variable factors as their degree of physical fitness and vitality, the regularity of their school attendance, the adequacy of their curriculum, and the competency of the instruction which they receive. The purpose of this study was to investigate the reading achievement of a representative group of fifty seventh grade students selected from a typical middle-class suburban community to see if such regular systematic reading progress had actually been made, and to discover, if possible, what factors had operated during the time these children had been in school to bring about the observed results.

Of the two hundred forty-four seventh grade children in the district, these fifty children had been in continuous attendance in this school system since the first grade. Their achievement in reading, therefore, was assumed by the authors to be an outgrowth of their total experience in these seven grades in this district.

Procedures

Each child's actual achievement was represented on an individual graph from grades one through seven on the basis of reading scores available for each school year. The expected achievement was also represented graphically on the basis of I.Q.'s obtained in the seventh grade. The study was augmented by a review of the cumulative records available for these fifty children.

Each I. Q. score used was taken from the California Short Form Test of Mental Maturity¹* which was given to each student in the 1956-57 school year. Through the use of a set of formulas developed by Horn², it was possible to compute the men-

tal age in months of each child at each grade level, and to convert this score into expected achievement in terms of grade placement for each child at each grade level.

To begin the study, all achievement data for each child were collected. The tests used were the same for each child, as they had been regularly given at approximately the same time each year to all pupils in the district. In grades one and two, the Ingram-Clark Diagnostic Test in Reading³ had been used, the score graphed having been derived from averaging the scores on the two parts of the test.

In grades three through seven, the California Achievement Tests in Reading⁴ were used, scores of the total reading tests being the ones graphed. In order to determine the extent of their reading and study skills, the Iowa Test of Basic Skills⁵ was also given to these seventh graders.

As a basis for the study, the following assumptions were made:

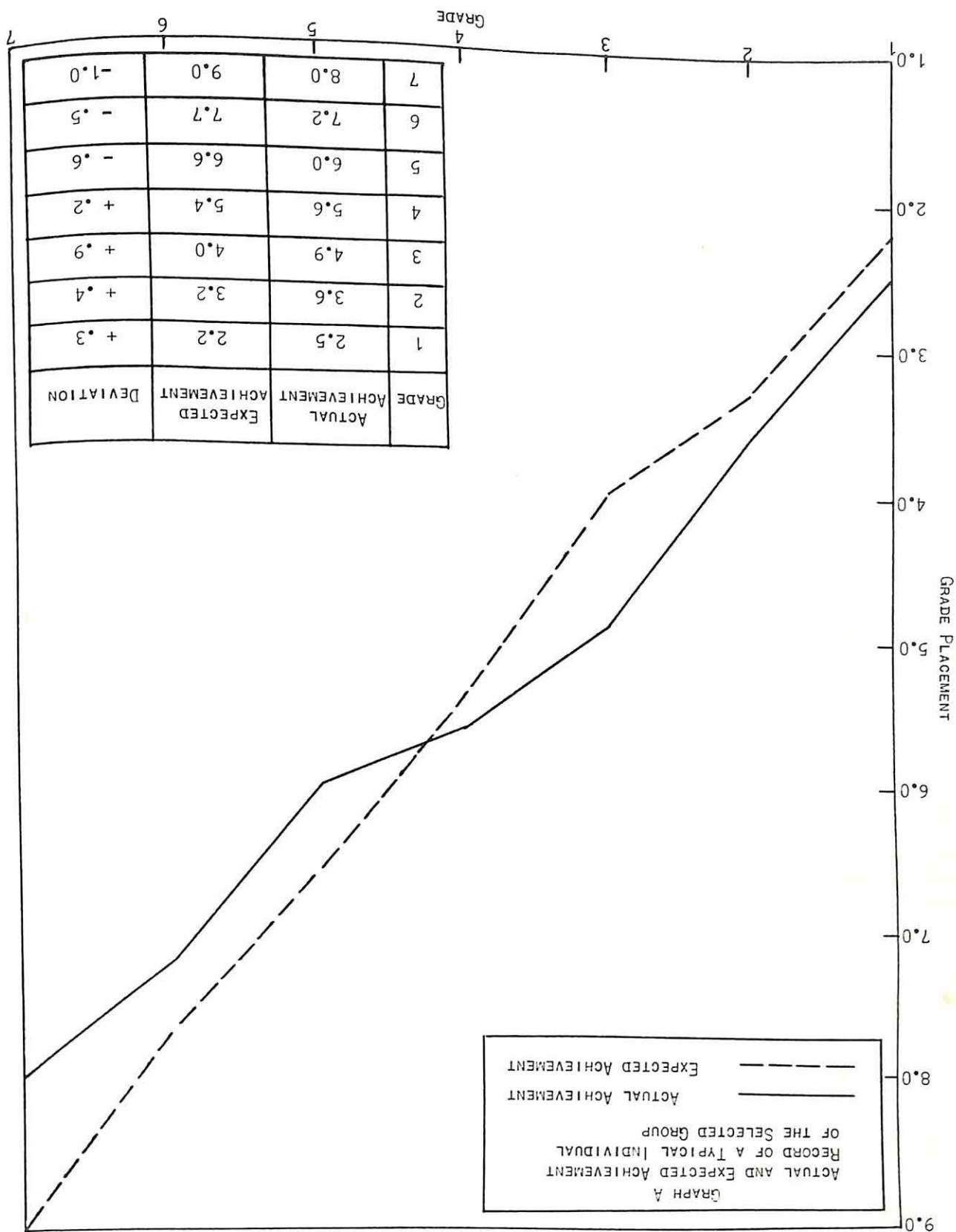
- 1) That the tests which were used actually measured what they purported to measure;
- 2) That each of the selected pupils applied himself equally well each time the tests were given;
- 3) That the use of grade placement scores would give satisfactory data for this study.

Findings

Following are six of the fifty-six graphs which were made. In all of the graphs, the dash line indicates the expected achievement in terms of grade level, while the solid line indicates the grade level actually achieved.

Graph A is a record of a typical individual of the selected group. As can be seen, the student worked above his expected achievement level until the fourth grade and then, although there is a small gain for the year, the achievement falls below the expected level in the fifth grade and does not come up to that level again.

* Footnotes will be found at the end of this article.



Graphs B and C compare boys and girls of the selected group. These are mean scores of the entire group of boys and the entire group of girls. The girls' scores seem slightly above the boys' scores, their first minus deviation appearing in the sixth grade while that of the boys appears first in the fifth grade.

Graphs D and E compare the highest third of the class, in terms of intelligence scores, with the lowest third of the class. Notice that although each classification begins at above the expected achievement level, the lower group shows the first minus deviation in the fifth grade while the higher group shows its first minus deviation in the sixth grade.

Graph F shows the mean result of the entire group, indicating that the group as a whole began above the expected achievement level, gradually attained an equal achievement level, and dropped below in the sixth grade. Although the total group made substantial gains in the seventh grade, it never quite attained the expected level.

Conclusions, Discussion, and Recommendations

When the reading achievement of these fifty students was compared with their expected achievement in terms of intelligence, as here measured, it was found that although for the first three or four grades the reading achievement and capacity paralleled consistently, at the fourth or fifth grades a minus deviation between actual achievement and capacity began to occur, and continued throughout the grades up to the seventh, with about eighty-five percent of the cases not returning to their expected achievement level.

From the results of this study it is evident that in the case of these fifty students studied in this school system, some causative factors relating to their lack of achievement must have been present in the fourth, fifth and sixth grades, for it is in those grades that the children did not work up to their capacity in the reading program. Some of these children may have had problems of health, both physical and mental, which would tend to lower their actual achievement, although a study of the cumulative records does not reveal any great problems of this sort. Some of the children may have been in classes which were too large to assure adequate instruction in reading, although in this district this does not appear to be the case.

No comparison was made in this study of the relative amount of time devoted to reading instruction in each of the grades. While variations in time spent in the teaching of reading skills would undoubtedly affect the results reported, it is the opinion of the authors that the most decisive factor in the minus deviation noted in reading achievement is the kind of instruction offered the children in the intermediate grades, as determined by personal observation of the authors and by informal interviews with the teachers involved in the study.

By these means it was possible for the authors

to ascertain some of the methods of reading instruction which had been used with these children in the intermediate grades. In one grade the children had been left entirely on their own in the reading program. Each child read from a different book and when he had finished, the teacher gave him a short comprehension quiz or asked comprehension questions such as which story he had liked best. The teacher provided no preparation for reading, no purpose for reading the story, no special vocabulary introduction, no discussion as to the main points or features in the pages covered.

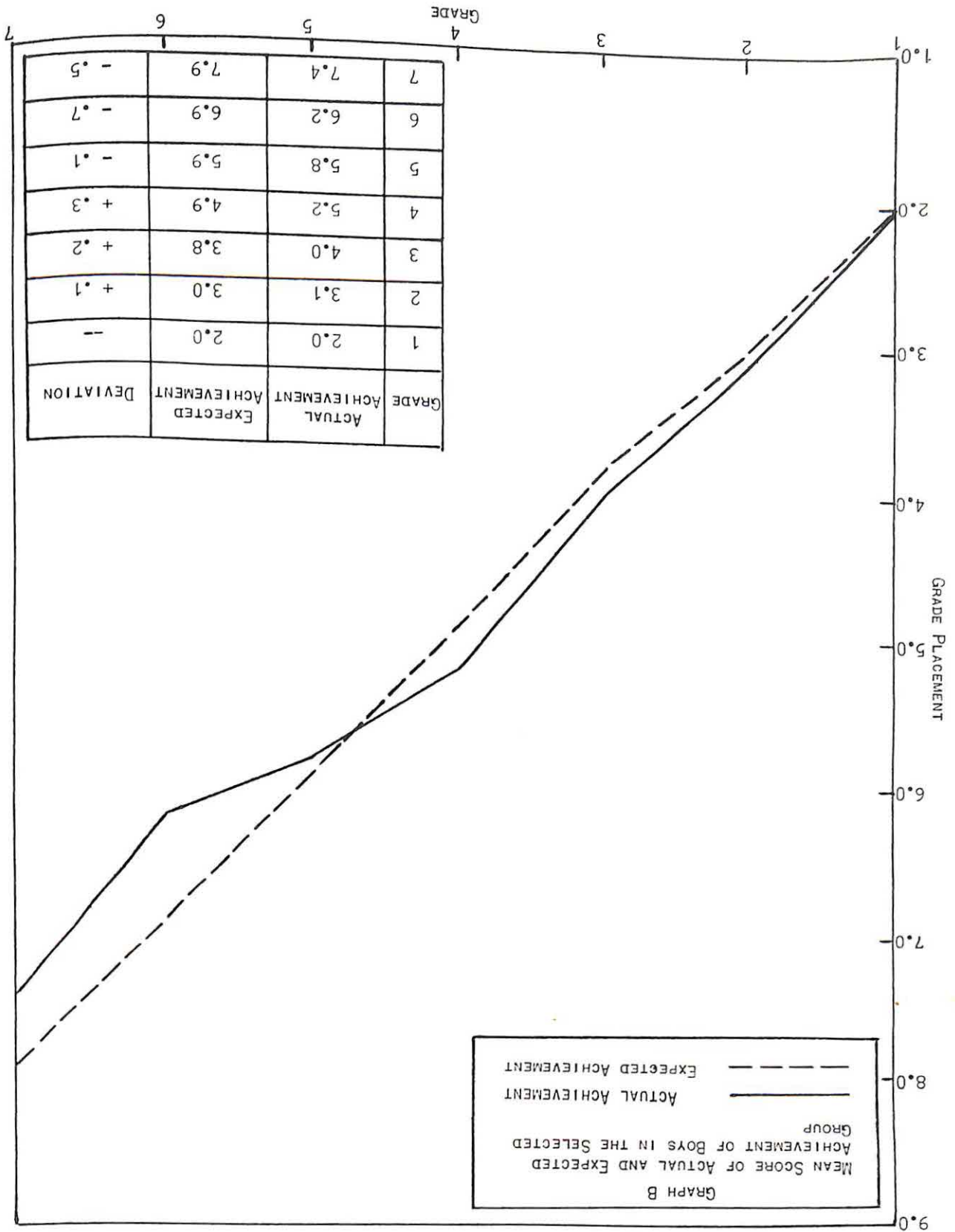
In another class situation, the teacher explained that the children sometimes had reading and at other times they didn't, the reading program being a somewhat spasmodic one. When there was to be a reading lesson, the teacher made an attempt to introduce the story, putting the unknown words on the board and pronouncing them. At the end of the period the teacher asked one or two questions as to the content of the story read.

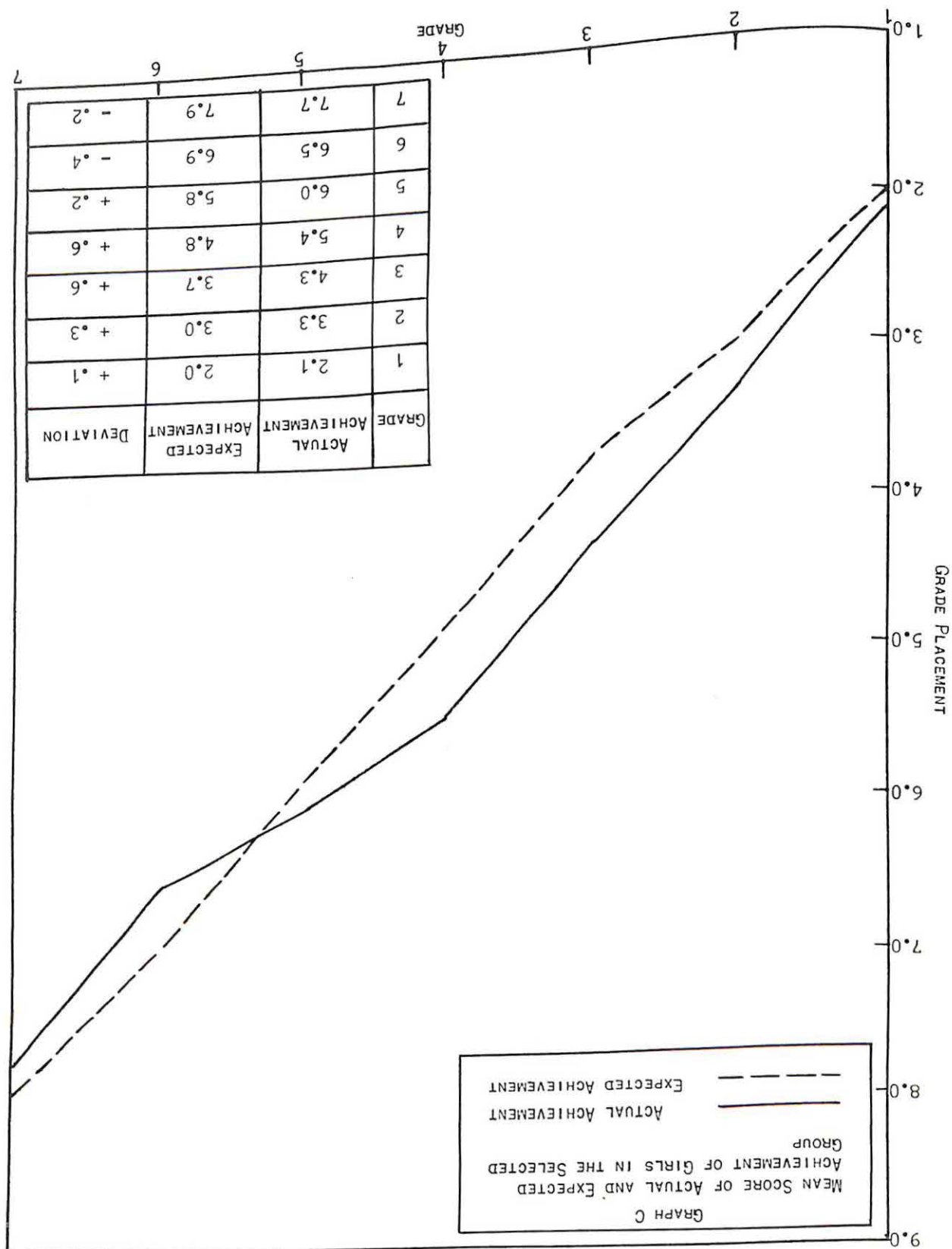
In still another class, the children read the assigned story silently, asking for help from the teacher if they encountered a strange word, and then went on to something else when finished. In many cases a story a day had been read, and the teacher stated that there just had never been time to do all of the other suggested extra work given in the teacher's manual. In some cases, the children were actually disturbed when a substitute teacher tried some of the enrichment activities with them. Evidently little use was made of the teachers' manuals in these classrooms.

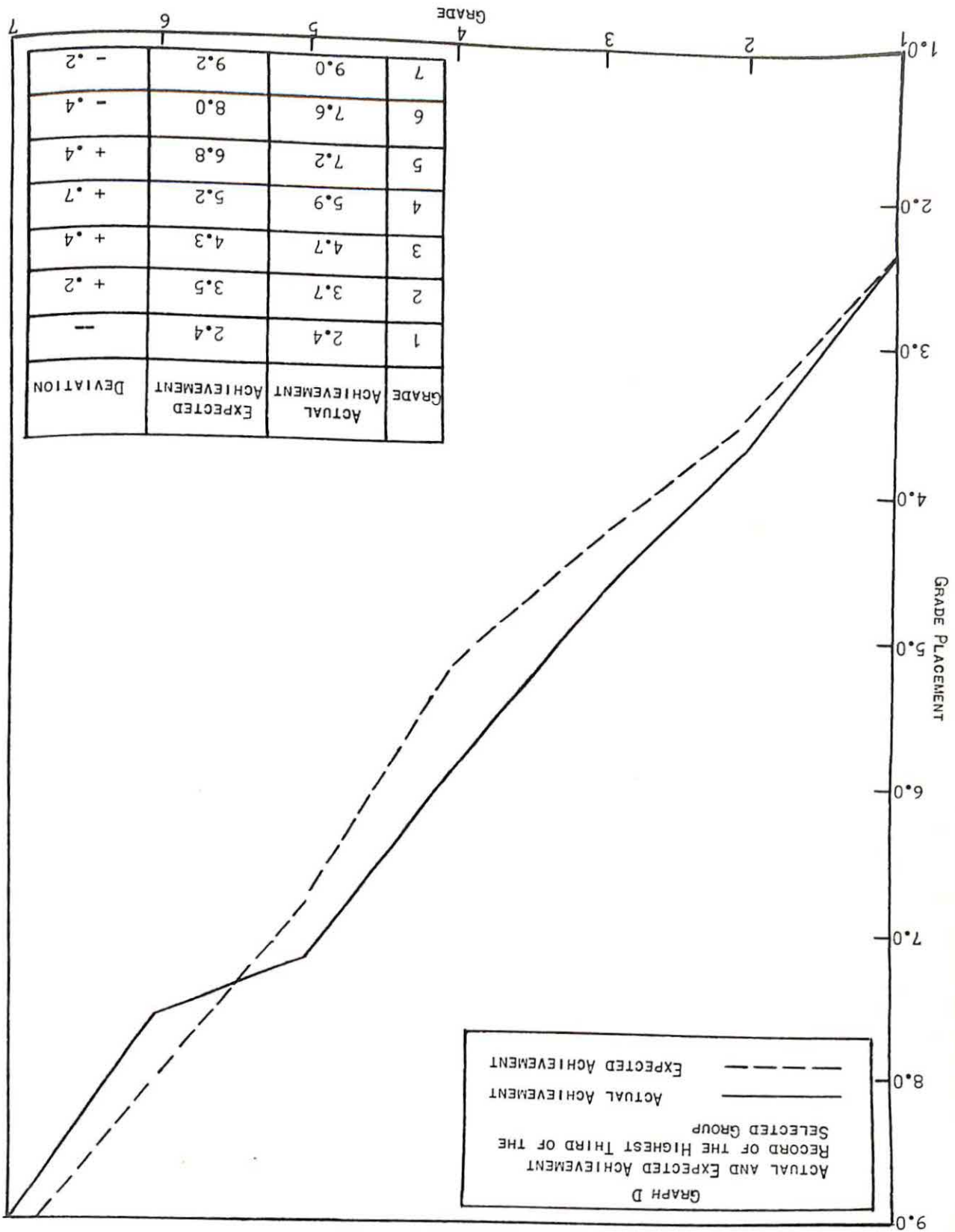
On the other hand, there were some teachers who had been carrying out a more adequate reading program. There were examples of purposeful reading activities going on in some classrooms and enrichment activities being enjoyed by groups of children--but unfortunately these seemed to be in the minority in the intermediate classrooms investigated.

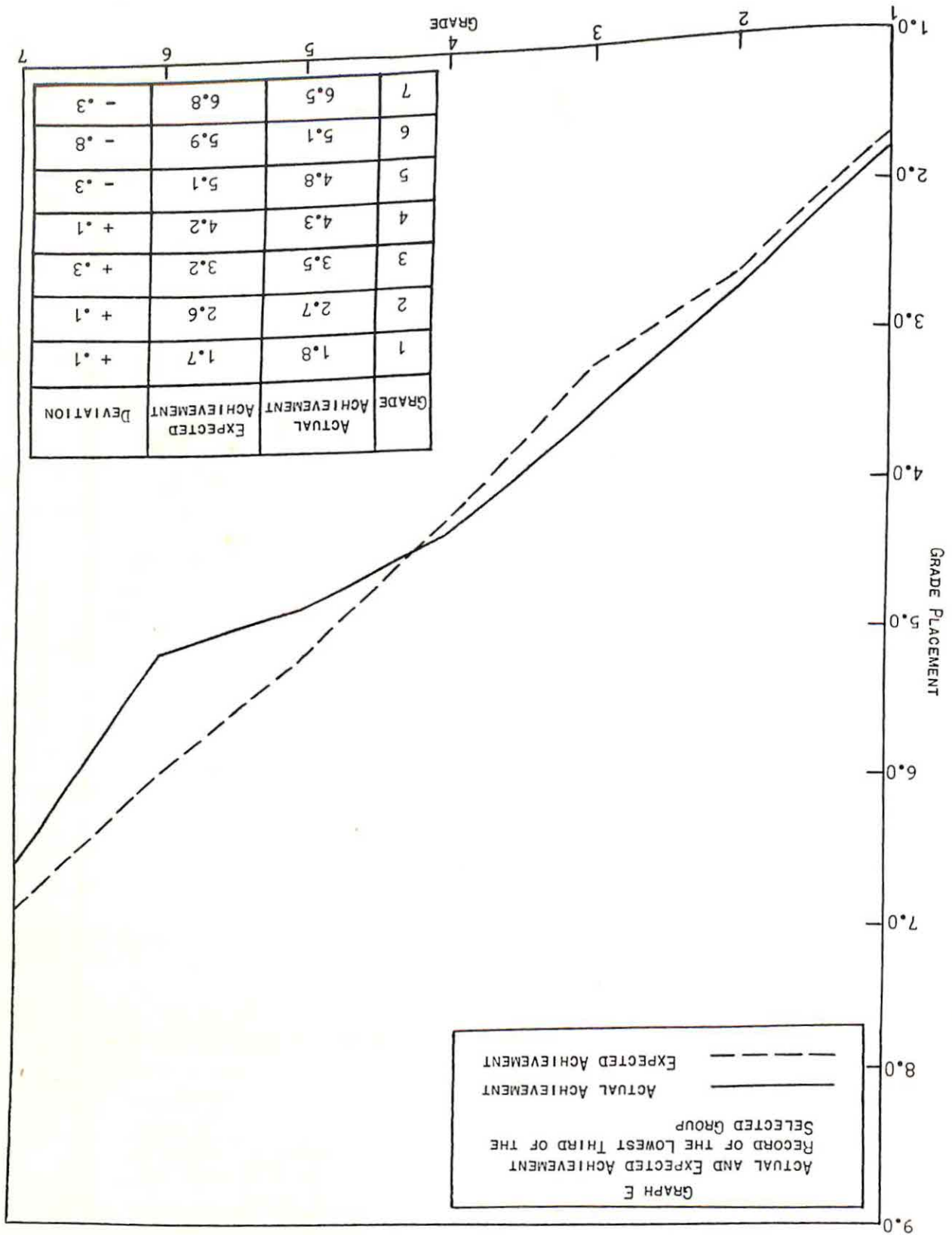
The research in this study does not make an attempt to evaluate middle grade teachers or methods, but rather to define sharply those areas in which expected progress had not been made, and to determine what methods had been used by the teachers in those areas. Further study of time allotments for reading instruction in the intermediate grades is indicated. The authors also feel that similar studies in comparable school districts would provide valuable information as to the extent of the problem.

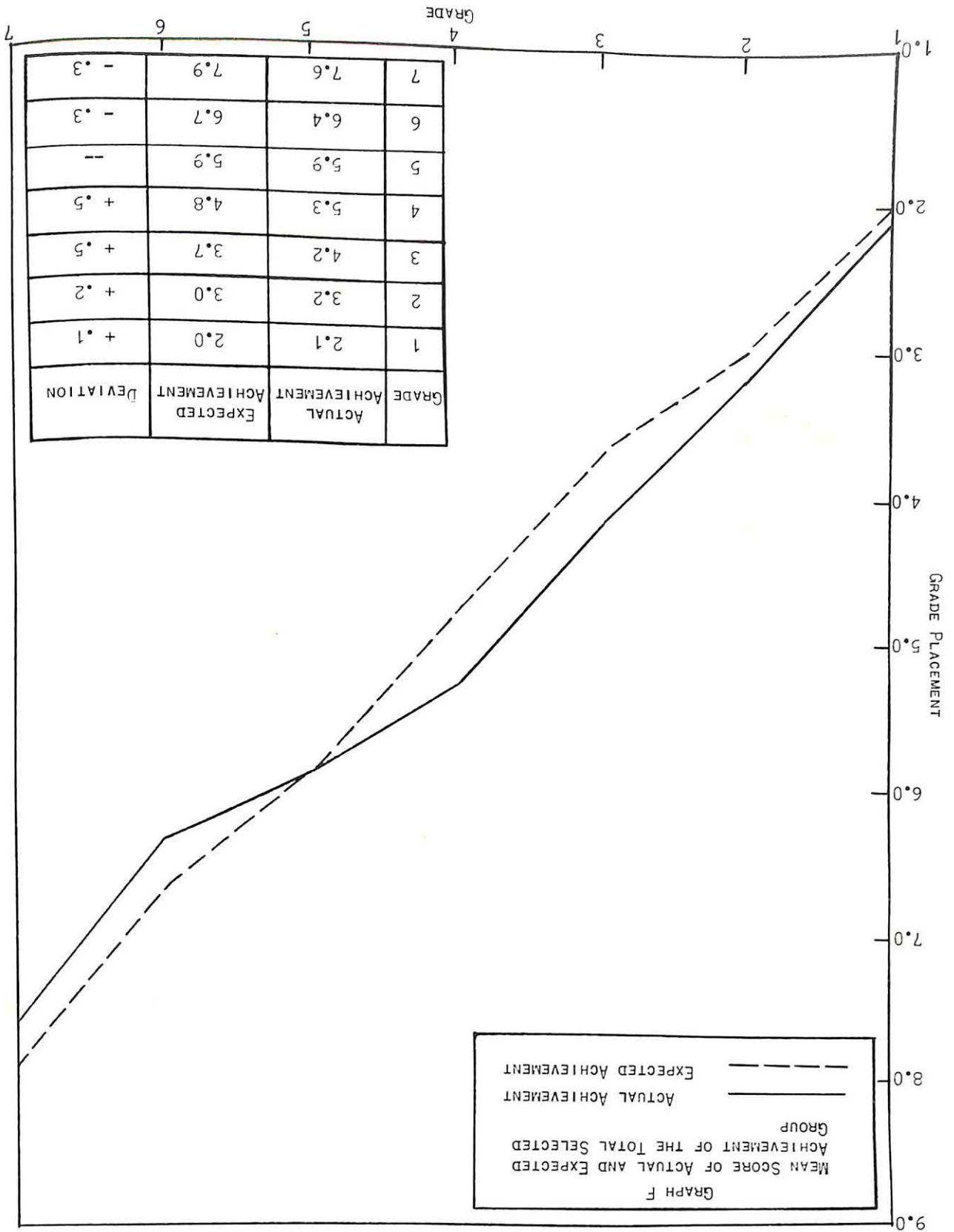
As a result of this study, recommendations have been made to the effect that teachers analyze their students' abilities early in the year and plan their reading program on the basis of each child's needs as well as the needs of the class as a whole. Further recommendations include the suggestion that teachers and children discuss the reading problems together and work out solutions as a team and that they set common purposes and goals. Moreover, provision must be made for enough time for instruction as well as for adequate, varied, and interesting











reading materials with which to work. Recommended procedure also includes consistent use of the teachers' manual, with teachers selectively choosing activities and exercises which are geared to the specific needs of children in their classes. It is vital, too, that intermediate grade teachers become aware of the necessity for teaching reading skills throughout the day, in all types of learning situations and with all kinds of subject matter, as well as during specially designated periods for reading instruction.

FOOTNOTES

1. Elizabeth T. Sullivan, Willis W. Clark, and Ernest W. Tiegs, California Test of Mental Maturity. Los Angeles: California Test Bureau, 1951.
2. Alice Horn, The Uneven Distribution of the Effects of Specific Factors. Southern California Education Monographs No. 12 (University of Southern California Press, 1941) and other unpublished research studies, Evaluation Section, Los Angeles City Schools.
3. Jessie E. Ingraham and Willis W. Clark, Ingraham-Clark Diagnostic Reading Tests, Parts I and II. Los Angeles: California Test Bureau, 1929.
4. Ernest W. Tiegs and Willis W. Clark, California Achievement Tests, Primary and Elementary Batteries. Los Angeles: California Test Bureau, 1951.
5. E. F. Lindquist and A. N. Hieronymous, Iowa Test of Basic Skills. Boston: Houghton Mifflin Company, 1955.

A COMPARATIVE STUDY OF ACHIEVING AND UNDERACHIEVING HIGH SCHOOL BOYS OF HIGH INTELLECTUAL ABILITY

EDWARD FRANKEL

Bronx High School of Science, New York

THIS STUDY WAS concerned with scholastic underachievement among intellectually superior high school students. The ever-broadening spectrum of our scientific and technological progress, from the harnessing of atomic energy to the conquest of outer space, has placed a special premium on talent and brainpower in all areas of human thought and endeavor. The young people whose scholastic performance lags far behind their intellectual ability represent a serious loss to society in terms of their potential contributions. In addition, failure to achieve at the level of their ability often leads to a depreciation of self-worth accompanied by unhappiness and frustration.

Design of the Experiment

The purpose of this study was to find some answers to the general question, "Why do students of seemingly similar high intellectual ability perform so differently academically?" This investigation proposed to study achieving and underachieving boys of the same high intellectual ability to determine possible causes for the differences in their academic performance. The areas explored for possible significant differences between the two groups were: 1) aptitudes, 2) interests, 3) personal problems, 4) health, 5) home and family background, 6) socio-economic status, 7) reaction to school subjects, 8) reaction to school, 9) out-of-school activities, 10) vocational and college planning and 11) academic performance in junior high school.

Instruments: The following instruments were used for gathering data related to each of these eleven areas--a) Differential Aptitude Tests, b) Kuder Vocational Preference Record, c) Mooney Problem Check List, d) School record, e) a Student Questionnaire of 39 items prepared by the investigator, and f) the Hamburger Scale for rating socio-

economic class.

The Experimental Group: The subjects participating in this study were selected from the male population of the senior class of June, 1957 at the Bronx High School of Science in New York City. The experimental group consisted of fifty pairs of boys, each pair composed of an achiever and an underachiever matched on the basis of equivalent I. Q., school entrance examination score, and age.

Definition of terms: Achiever was defined as a student in the top or first quartile of his class with a scholastic average of at least 89 per cent for the tenth and eleventh years. Underachiever was defined as a student in lowest or fourth quartile of the same class with a scholastic average of 79 per cent or less.

Academic Environment of the Study: The study was limited to the Bronx High School of Science because not only was it an ideal source of high ability subjects for this study but also because the investigator has been a teacher and guidance counselor at the school since 1940. The school, which came into being in 1938, was designed to meet the needs of high ability students interested in science and mathematics. The school population is about 2400 one third of whom are girls. About 800 students are admitted annually, 150 from the elementary schools to the ninth year and 650 from the junior high schools to the tenth year. About 98 per cent of those who are graduated from the school enter college. The school has been described by Wolfe "as being the outstanding exception in this country in providing stimulation and training for bright youngsters with scientific interests."

About four to five times as many students as can be accommodated apply for admission to the school. A program for selecting students has evolved which includes a written examination administered at the school. It is an objective test consisting of two

* This paper is based upon the author's doctoral study entitled "A Comparative Study of Achieving and Underachieving High School Boys of High Intellectual Ability" submitted in partial fulfillment of the requirements for the Ph. D. degree at the Graduate School of Education, Yeshiva University, granted in June, 1958.

parts: a) English, which includes reading comprehension and vocabulary, and b) Arithmetic. Ninety minutes are allowed for the entire test.

The curriculum of the school aims to prepare students of high ability to meet the admission requirements of liberal arts colleges as well as engineering and technical schools. The subjects are those usually offered in an academic high school. However, the curriculum is enriched by and supplemented with a broad program of elective courses in science and mathematics. In addition, opportunities for acceleration and advanced study are possible by offerings of college-level courses in English, Mathematics, Biology, Chemistry, and Physics.

Selection of the Experimental Group

Criteria for Selection: In matching an achiever with an underachiever, a maximum difference of five points in I. Q., five points in school entrance examination score, and twelve months in age, were used as a basis for obtaining equivalent groups.

The scholastic average which was used to define achiever and underachiever was obtained by calculating the mathematical average of the final grades received in all major subjects during the tenth and eleventh year. Ninth year subjects were not included in the calculations since eighty percent of the class were in junior high school during that year. Since one of the constant factors in this study was the school environment, marks received in other schools where differences in standards and marking systems might exist, would tend to introduce a variable not included in this study. To obtain equivalent groups optimally matched, only white boys were selected for this study. In addition, ninth grade entrants were matched only with ninth grade entrants, and those who came from junior high school and entered school in the tenth grade were matched with tenth grade entrants.

Testing Criteria by Preliminary Study: In May, 1956, a preliminary study, involving the class of June, 1956, was undertaken to test the criteria. It was found that forty-two pairs of boys could be matched, using the criteria selected.

Criteria Characteristics of the Experimental Group: In September, 1956, using the criteria established for matching, it was possible to select a maximum of fifty pairs of boys from the class of June, 1957, which had a male population of 468. (Table I)

The t-values for I. Q., entrance examination score, and age indicated that there were no significant differences between the two groups for these criteria. On the other hand, the t-value of 23.76 for scholastic average clearly showed that the groups were significantly different in academic achievement, the mean difference being 17.1, with a minimum of 9.5 and a maximum of 37.2.

The entrance examination test score is a composite of the English and the Arithmetic parts. In order to avoid inequalities in matching which might be masked by the composite score, the two groups

were compared for the English and Arithmetic scores separately. No significant differences were found between the two groups in English and Arithmetic parts of the entrance examination. The two groups were matched not only for the composite entrance score but also for the English and Arithmetic parts.

Findings

Aptitudes: The results which were obtained by the fifty pairs of boys on the four Differential Aptitude Tests which were administered in October, 1956, are shown in Table II.

The Numerical Ability test discriminated most sharply between the two groups with a highly significant t-value of 5.15. The achievers were definitely superior in their ability to understand numerical relationships and in handling numerical concepts.

In addition, the achievers showed definite superiority in Verbal Reasoning, the ability to understand concepts framed in words, the t-value being 2.83. Together these two tests are regarded as a measure of general learning ability. The Space Relations and the Abstract Reasoning tests revealed no significant differences between the groups.

The results of the Scholastic Aptitude Tests for May, 1956 and January, 1957 and the Scholarship Qualifying Tests for April, 1956, which half the subjects had taken, substantiated the DAT findings. The achievers showed significantly greater aptitudes than the underachievers in the verbal and mathematical areas.

Interests: A comparison of the interests of the two groups as measured by the Kuder Vocational Preference Record administered in October, 1956 revealed the data shown in Table III.

The interests of the U's were significantly greater in the mechanical area whereas those of the A's were in the scientific, the t-values being significant at the 01 level, 3.18 for the former and 3.01 for the latter. In addition, the A's were more interested in the computational and the U's in the artistic, the t-values being significant at the 05 level. The other six areas showed no significant differences.

Personal Problems: The Mooney Problem Check List, administered in September, 1956, which was used as a measure of the personal problems of the two groups, showed no statistically significant difference in the total number of problems underscored although the A's underscored 723 and the U's 906 problems. An analysis by areas of the number of problems underscored is shown in Table IV.

"School" was the only area in which the U's presented significantly more problems than the A's. There were no significant differences in the other six areas.

Of the 210 items of the Mooney, differences in the frequency of responses of the A's and U's significant at and beyond the 05 level occurred for only ten items. Of these, nine were underscored by significantly more U's than A's. Eight of these

TABLE I
A COMPARISON OF CRITERIA FOR THE EXPERIMENTAL GROUP: FIFTY
PAIRS OF ACHIEVING AND UNDERACHIEVING BOYS SELECTED
FROM THE CLASS OF JUNE, 1957

Criteria	Group ^a	Mean	S. D. ^b	Range	t-Value
Intelligence Quotient	A	140.66	12.05	114-164	0.80
	U	141.00	11.79	115-164	
Entrance Exam. Test Score	A	161.70	13.60	137-190	0.21
	U	161.80	12.27	142-187	
Age in Years	A	17.00	0.60	16.00-18.34	0.03
	U	17.00	0.55	15.83-18.34	
Scholastic Average	A	91.24	1.69	89.0-96.2	23.76**
	U	74.13	4.07	59.0-79.5	

a The letters "A" and "U" are used to identify the achievers and underachievers respectively.

b Standard Deviation.

** Significant at or beyond the 01 level.

TABLE II
RESULTS OF DAT BATTERY TAKEN BY EXPERIMENTAL GROUP

DAT Test	Group	Mean	S. D.	Range	t-Value
Space Relations	A	61.48	16.73	28-92	0.43
	U	62.93	18.90	26-99	
Abstract Reasoning	A	40.16	4.44	25-47	0.75
	U	39.36	4.86	25-46	
Verbal Reasoning	A	42.40	4.66	29-49	2.83**
	U	39.85	4.98	31-50	
Numerical Ability	A	35.54	2.75	28.75-40.00	5.15**
	U	31.95	4.48	16.00-37.50	

** Significant at and beyond the 01 level.

TABLE III
COMPARISON OF AREAS OF INTERESTS OF EXPERIMENTAL GROUP AS
MEASURED BY THE KUDER VOCATIONAL PREFERENCE TEST

Area of Interest	Group	Mean	S. D.	Range	t-Value
Outdoor	A	41.08	14.78	11-69	0.31
	U	40.04	16.88	8-68	
Mechanical	A	35.58	15.51	2-66	3.18**
	U	43.42	12.19	11-63	
Computational	A	33.26	9.42	11-50	2.39**
	U	29.00	8.47	15-50	
Scientific	A	54.76	12.50	20-70	3.01**
	U	45.64	13.60	15-66	
Persuasive	A	38.20	13.41	11-69	1.88
	U	42.84	13.40	21-72	
Artistic	A	20.38	9.19	6-42	2.08*
	U	24.88	11.38	7-48	
Literary	A	24.78	8.69	6-39	1.32
	U	22.96	7.83	4-36	
Musical	A	13.04	6.06	4-25	1.29
	U	11.12	7.81	3-29	
Social Service	A	38.94	12.70	18-71	1.55
	U	34.92	13.29	12-62	
Clerical	A	41.04	14.74	14-77	0.39
	U	42.18	12.97	16-80	

** Significant at and beyond the 01 level.

* Significant at and beyond the 05 level.

TABLE IV

DISTRIBUTION OF PROBLEMS BY AREAS UNDERScoreD ON THE MOONEY
PROBLEM CHECK LIST BY ACHIEVING AND UNDERACHIEVING BOYS

Area	Group	Mean	S. D.	Range	t-Value
Health and Physi- cal Development	A	2.22	2.03	0-9	0.89
	U	1.84	1.85	0-7	
School	A	1.64	1.75	0-7	6.62**
	U	5.28	3.14	0-14	
Home and Family	A	1.74	2.60	0-12	0.21
	U	1.88	2.63	0-12	
Money, Work and Future	A	3.32	2.94	0-15	0.04
	U	3.30	2.75	0-11	
Boy-Girl Relations	A	1.96	2.56	0-13	0.16
	U	1.88	2.43	0-11	
Relations to People	A	1.30	1.51	0-5	0.42
	U	1.46	2.41	0-11	
Self-Centered Concerns	A	2.28	2.46	0-12	0.46
	U	2.50	2.85	0-12	

** Significant at and beyond the 01 level.

TABLE V

VOCATIONAL CHOICES OF ACHIEVERS AND UNDERACHIEVERS

Group	Scientific		Non Scientific	Undecided
	Tech.	General		
A	14	22	0	14
U	22	9	4	15
Chi-square	1.36	4.64*	2.25	0

* Significant at and beyond the 05 level.

were concerned with "School" and the other one was "being stubborn". "Family Quarrels" was the only item selected by more A's than U's. In answer to the free question "What problems are troubling you most?", the U's reiterated their concern with school and marks, whereas the A's were interested chiefly in the choice of college and vocation.

Health: The information gathered by the Student Questionnaire regarding health showed no differences between the two groups in weight, height, hearing, speech, general state of health, and physical disabilities. The U's, however, reported significantly more days absent from school for health reasons, the t-values being 2.34. The A's, on the other hand, registered significantly more health complaints, chiefly acne and allergies; the t-value being 2.47.

Family and Home Background: Questionnaire responses indicated no differences between the two groups with respect to a) number of rooms in the home, b) number of people living at home, c) size of family, d) number of disrupted family patterns, and e) birth order of the subjects. Differences were found, however, in the education and occupation of the parents. Using the Edward's Scale for the classification of occupations, more of the fathers of the A's than U's were found in the top three groups--1) professional, 2) semi-professional, and 3) proprietors, managers, and officials. With respect to number of years of schooling completed by parents of the two groups, it was found that the fathers of the A's had significantly more formal education than the mothers, the t-value being 2.33. No significant difference between the fathers and mothers of the U's in this respect was found; they had about the same amount of formal education. Significantly more working mothers were reported by the U's than A's, twenty-nine to seventeen, the chi-square value for the difference being 6.43 which is significant at and beyond the 05 level. More of the mothers of the U's than A's were in the lower three occupational groups. Almost two-thirds of them were typists, bookkeepers, secretaries, and saleswomen.

Socio-Economic Status: Using the Hamburger Scale for determining socio-economic level, the A's came from families which were rated significantly higher than those of the U's.

Reaction to School Subjects: In response to the Questionnaire items requiring the groups to name the major school subjects a) liked best, b) found easiest, c) liked least, and d) found most difficult, the following reactions were gleaned: The A's ranked mathematics as the easiest, and mathematics and science as the best liked, English as the most difficult and least liked. The underachievers, on the other hand, selected science as the easiest and best liked subject, and foreign language as the most difficult and least liked. Significantly more A's than U's selected mathematics as the easiest and best liked subject, and English as the most dif-

ficult. In contrast, the U's chose science as the easiest and mathematics as the most difficult school subject.

Reaction to School: The criteria selected to measure reaction of the two groups to school were a) attendance, b) deportment, c) participation in extra-curricular activities. These data were obtained from official school records. According to official attendance records, the U's were absent from school significantly more often than the A's, the t-value being 3.58, significant at the 01 level. School discipline records showed about four times as many offenses recorded against U's than A's, the t-value being 4.07, significant at and beyond the 01 level. With respect to extra-curricular activities, the A's engaged in significantly more, the t-value being 5.35, significant at the 01 level. Twelve U's had records of no extra-curricular activities. The A's engaged significantly more frequently in students government and publications, in science and mathematics clubs, and in the Social Studies club.

Out-of-School and Leisure Time Activities: The U's belonged to significantly more athletic and social clubs and were somewhat more interested in the Scout movement. Insofar as leisure time activities were concerned, the A's tended to spend more time reading whereas the U's were more interested in shop-work activities.

Vocational and College Planning: Questionnaire responses with respect to vocational planning are shown in Table V. Significantly more A's planned to enter general fields of science such as mathematics, biology, chemistry and physics. The U's, in contrast, tended more in the direction of applied sciences and technical fields. Responses to a Guidance Committee questionnaire, in June, 1957, at the time of graduation, dealing with the choice of college program appear in Table VI.

Significantly more A's than U's planned to follow a liberal arts college program majoring in science, whereas more U's than A's expected to enter non-science fields such as business administration, accountancy, and the like. Of those who were undecided, as seen in Table V, more U's than A's selected non-science fields.

The Questionnaire also revealed that more U's than A's, as evidenced by the chi-square value of 4.11, significant at the 05 level, expected their parents to finance completely their college education.

Junior High School Record: The records of forty-three pairs of boys who entered the school in the tenth year from junior high school were studied to determine whether their patterns of academic achievement was developed in the senior high school or had previously been established in the junior high school. It was found that twenty-eight achievers and thirty underachievers had been regarded as intellectually gifted and had been placed in Special Progress (S. P.) classes which had completed the three years of junior high school in two years.

TABLE VI
CHOICE OF COLLEGE PROGRAM BY THE EXPERIMENTAL GROUP

Group	Technical	Liberal Arts		Non-Liberal Arts	Undecided
		Science	Non-Science		
A	14	27	5	0	4
U	18	12	7	8	5
Chi-square	1.47	5.03*	0.08	6.13*	0

* Significant at and beyond the 05 level.

TABLE VII
SCHOLASTIC AVERAGE FOR THE NINTH YEAR OF JUNIOR HIGH SCHOOL
OF FORTY-THREE PAIRS OF ACHIEVERS AND UNDERACHIEVERS

Group	Mean	S. D.	Range	t-Value
A	90.98	3.00	97.00-84.00	
B	82.44	5.08	91.20-71.00	10.33**

** Significant at and behond the 01 level.

In addition, the results of the reading and arithmetic achievement tests, which had been taken in the ninth year of junior high school, attested to their superiority. The Stanford Reading Test showed no differences between the groups. Approximately one third of each group made perfect scores and practically the entire group made grade equivalent scores above 10.0, a value derived by extrapolation. The mean grade equivalent score for the entire group was approximately 11 plus. On the other hand, on the New York Arithmetic Test, the equivalent scores of the A's was higher than those of the U's, the t-value being 3.12, significant at the 0.01 level. Nine A's made perfect scores, and practically the entire group scoring above 9.4 also derived by extrapolation. The mean grade equivalent score of the A's was 11.59 and that of the U's was 10.84.

A comparison of the school grades received by the two groups in the ninth year of junior high school showed results set down in Table VII.

The scholastic record of the two groups for the ninth year indicated that they were performing distinctly differently. Three quarters of the A's attained averages of 89 per cent or better, the lowest being 84. Half of the U's earned less than 80 per cent, fifteen were between 80 and 84, and only four were above 89 per cent. Three quarters of the U's had averages of less than 84 per cent.

Conclusions

1. Aptitudes: Although the pairs were matched for equivalent I. Q. and school entrance examination, the achievers proved to be distinctly superior to the underachievers in mathematical and verbal aptitudes, particularly in the former.

2. Interests: The interest patterns of the two groups were distinctly different. The interests of the achievers were greater in mathematics and science whereas those of the underachievers were in the mechanical and artistic areas.

3. Personal Problems: While the chief concern of the underachiever appeared to be his present scholastic inadequacies, the achiever's was primarily thinking about the future, college and vocational choices.

4. Health: Although the U's reported twice as many days absent from school for health reasons, and official attendance records disclosed that the U's were absent from school significantly more frequently, they registered fewer specific health complaints on the Student Questionnaire, and under-scored fewer items in the Health and Physical Development area of the Mooney. No evidence was found to lead one to believe that the two groups differed significantly in health. It seemed likely that the more frequent absence from school for health reasons reported by the U's might not necessarily have been the result of physical illness.

5. Home and Family Background: Although the physical aspects of the families of the two groups

were very much alike, significant differences in the education and occupation of the parents existed. More of the fathers of the A's were in the top three occupational groups and they had more formal schooling than their wives. More of the mothers of the U's were working and they had at least as much schooling as their husbands.

6. Socio-Economic Status: As expected, the families of the A's were rated higher on the Hamburger Socio-Economic scale.

7. Reaction to School Subjects: The selection by the achievers of mathematics as the easiest, and science and mathematics as the best liked school subjects, was probably a reflection of their superior aptitude and greater interest in these areas. Similarly, the distaste of the U's for mathematics mirrored the difficulty which they had with this subject.

The negative reactions of the A's to English might be an expression of their science-mathematics preference. The selection of science as the easiest and best liked subject by the U's might be explained by the fact that the sciences with their concomitant laboratory opportunities offered an outlet for their mechanical interests.

8. Reaction to School: It was not surprising to find that the U's evidenced negative attitudes toward school, the major source of their personal problems, in terms of poorer attendance records, more recorded disciplinary offenses, and participation in fewer extra-curricular activities. In general, the U's were more recalcitrant, less conforming, and less happy at school. The achievers, in contrast, were more conforming, rarely broke school regulations, participated in more school activities and assumed positions of leadership and responsibility.

9. Out-of-School and Leisure Activities: The greater participation of the U's in out-of-school organizations such as Social and Athletic clubs, and the Scouts, may have been a substitute for school activities.

10. Vocational and College Planning: The achievers appeared to regard college as preparatory for a career in science with the expectation of going on to graduate school for specialization. The underachievers tended to think of college in terms of direct vocational preparation; those going into the sciences were planning careers in applied and technical fields. However, a substantial number of U's planned to prepare for and enter non-science fields.

11. Junior High School Academic Performance: Notwithstanding the superior intellectual ability of the two groups, the ninth year junior high school record left little doubt that the two groups performed differently in terms of academic achievement. In general, the achievers maintained their high scholastic record while the performance of the underachievers deteriorated. The difference in the mean scholastic average of the two groups was twice as great in high school as in junior high school. It appeared probable that the factors relating to scholastic underachievement of this group may have been operating before these students entered the high

school.

REFERENCES

I. Books

1. Brandwein, Paul F., The Gifted Student as Future Scientist. New York: Harcourt, Brace and Company, 1955.
2. Bray, Douglass Weston, Issues in the Study of Talent. New York: King's Press, Columbia University, 1954.
3. Cole, Charles C., Jr., Encouraging Scientific Talent. New York: College Entrance Board, 1956.
4. Havighurst, R. J.; Stivers, E. and De Haas, R. F., A Survey of the Education of Gifted Children. University of Chicago Press, November, 1955.
5. Roe, Anne, The Making of a Scientist. New York: Dodd, Mead and Company, 1953.
6. Terman, Lewis and Oden, M. H., The Gifted Child Grows Up. Genetic Study of Genius, Vol. IV. Stanford: Stanford University Press, 1947.
7. Witty, Paul (ed). The Gifted Child. New York: Heath and Company, 1951.

II. Periodicals

1. Armstrong, Marion E., A Comparison of the Interests and Social Adjustments of the Under-Achievers at the Secondary School Level, Doctor's Thesis, University of Connecticut, 1955.
2. Beasley, Jane, Underachievement: Review of Literature: Teachers' College, Columbia University, March, 1957 (mimeographed).
3. Gowan, John C., "The Under Achieving Gifted Child." Exceptional Children, XXI (April, 1955), 247-249.
4. Gowan, John C., Underachievement of Gifted Children. Los Angeles State College (mimeographed).
5. Gough, Harrison G., "Factors Related to the Academic Achievement of High School Students." Journal of Educational Research, XL, (1949) 65-78.
6. "Potential Use of Personality Tests in School and Colleges". Fifth Annual Western Regional Conference on Testing Problems (April 13, 1956) Los Angeles: Educational Testing Service, 2-19.
7. Hamburger, Martin, A Revised Occupational Scale for Rating Socio-Economic Class. Teachers' College, Columbia University, May, 1957 (mimeographed).
8. Hieronymus, A. N., "A Study of Social Class Motivation: Relationship Between Anxiety for Education and Certain Socio-Economic and Intellectual Variables". Journal of Educational Psychology, XLII (1951) 193-205.
9. Kimball, B., "Completion Technique in a Study of Scholastic Underachievement". Journal of

- Consulting Psychology, XVI (October, 1952) 353-358.
10. Layton, E. T., A Study of the Factors Associated with Failure in the Ninth Grade at Hempstead High School. Doctor's thesis, New York University, 1951.
11. MacCurdy, Robert D., Characteristics of Superior Science Students and Some Factors that Were Found in Their Background. Doctor's Thesis, Boston University, 1954.
12. MacCurdy, Robert D. "Characteristics of Superior Science Students and Their Own Sub-Groups." Science Education, XL (February 1956). 3-24.
13. Malpass, L. F., "Some Relationships Between Students' Perception of School and Their Achievement". Journal of Educational Psychology, XLIV (December, 1953) 475-482.
14. McQuary, John P., "Some Relationships Between Non-Intellectual Characteristics and Academic Achievement". Journal of Educational Psychology, XLIV (April, 1953) 215-228.
15. Morgan, H. H., "A Psychometric Comparison of Achieving and Non-Achieving College Students of High Ability", Journal of Consulting Psychology, XVI (August, 1952) 292-298.
16. Nason, Leslie J., Patterns of Circumstances Related to Education Achievement of High School Pupils of Superior Ability. Unpublished Ph.D. thesis, University of Southern California, 1954.
17. Portland Public Schools, A Report Summarizing Four Years of Progress by the Cooperative Program for Students of Exceptional Talent. Portland, Oregon, March, 1957 (mimeographed).
18. Ratchick, Irving, Achievement and Capacity: Study of Pupils with Low Achievement and High Intelligence Quotient with Pupils of High Achievement and High Intelligence Quotients in a Selected New York City High School. Doctor's Thesis, New York University, 1953.
19. Ricciuti, N. N. and Sadacca, Robert, The Production of Academic Grades With a Projective Test of Achievement Motivation: II Gross Validation at the High School Level. Princeton: Educational Testing Service, June 6, 1955.
20. Roe, Anne, "A Psychologist Examines Sixty-Four Eminent Scientists". Scientific American, CLXXXVII, No. 5 (November, 1952), 21-25.
21. Talented Youth Project, Student Attitude Toward School Inventory. Evanston, Illinois, Horace Mann-Lincoln Institute of School Experimentation, Teachers' College, Columbia University, May 13, 1957.
22. Terman, Lewis, "The Discovery and Encouragement of Exceptional Talent", American Psychologist, IX (June, 1954) 221-230.
23. Wolfle, Dael, "Intellectual Resources". Scientific American, CLXXXV, No. 2, (September, 1951) 42-46.

AN EMPIRICAL STUDY OF PERFORMANCE IN MATHEMATICS AND PERFORMANCE ON SELECTED ENTRANCE EXAMINATIONS

JOSEPH R. MARCHES
University of Maryland

THE INVESTIGATION here reported seeks answers to two questions: To what extent can the predictive qualities of selected college entrance examinations serve counselors and advisors in guiding education majors who are required to take course work in mathematics as a part of their curriculum? Assuming the validity and reliability of the scores from the selected tests, to what extent do such variables as motivation and interest intervene to reduce the predictability value of these instruments (4)?

The first question can be answered by testing the relationship between the entrance test scores of students majoring in education and the letter grades they received in mathematics course work.^{1*} Theoretically, a measure of high relationship indicates a high degree of predictability for the test and vice versa. Counselors and advisors could determine the relative risk in assigning mathematics course work on the basis of entrance test score rank. A "high" rank score would suggest high probability of success while a "low" rank score would suggest some need for preliminary work in a preparatory course or special tutoring in the foundation of general mathematics.²

This, however, is a simplified theoretical explanation in the prediction of performance. As for what occurs in the empirical setting, a corollary to this theoretical framework must be entertained. This is related to the problem of answering the second question. Theoretically, motivational and interest factors intervene to a greater or lesser degree to produce performance levels not totally congruent with entrance test score rank (1).

Obviously, theory serves to explain the disparity between entrance test score rank and actual performance but it cannot fully answer the question as to the degree to which intervening motivational and interest variables are operative for individual cases. Herein lies the problem of performance prediction. The sample tested in this study is primarily

composed of students majoring in elementary education although mathematics education and industrial education majors are well represented. The sample is segmental in nature and is composed of students whose entrance dates range from 1950 through 1955. The testing program was virtually unchanged during this period of time. The cases were recorded without reference to age, sex, or marital status. An attempt was made to minimize the inclusion of transfer students from other colleges within the University and to exclude transfer cases from schools outside of the University. The total number of cases employed in the sample is 100. However, since most students enrolled in more than one mathematics course, the N value is greater than 100 for most of the cross-tabulations.

Tables I, II, and III show percentage distributions of mathematics course grades and percentile scores for selected entrance examinations. The overall pattern appears to be absence of a gradient for percentage values consistent with the gradient for percentile rank scores.

Rank Correlations (3). Several rank correlation values were computed to test the relationship between the observed performance and the expected performance in mathematics on the basis of entrance tests. The disparity between the actual and the theoretical conditions is demonstrated schematically with corresponding correlation values. The greater the disparity, the greater the support for the theory of intervening variables. The reverse is true when the disparity is slight. It is interesting to note that the modal or most popular grade achieved in mathematics is "C" for most of the entrance tests regardless of percentile score range. This is in spite of the fact that Tables I, II, and III show higher percentages of "A" and "B" grades for the upper ranges of test scores.

The Performance Profile. A demonstration of the relationship between performance in specific mathematics courses most frequently taken by ed-

* Footnotes will be found at the end of the article.

TABLE I

PERCENTAGE DISTRIBUTION OF GRADES ACHIEVED IN ALL MATHEMATICS COURSES ENROLLED FOR EACH GIVEN RANGE OF TOTAL AND COMPONENT PERCENTILE SCORES OF THE A.C.E.

Range of Percentile Scores	Total per cent for each component	Percentages														
		A			B			C			D			F		
		Tot.	Q	L	Tot.	Q	L	Tot.	Q	L	Tot.	Q	L	Tot.	Q	L
80-100	100	16	21	22	28	29	28	33	29	31	13	11	16	10	10	3
60-79	100	17	4	13	28	33	23	33	39	35	15	17	19	7	7	10
40-59	100	2	15	2	14	12	11	44	41	36	32	22	26	8	10	25
20-39	100	6	7	5	14	16	22	32	35	41	30	35	25	18	7	7
0-19	100	7	2	8	7	3	8	49	40	40	30	26	32	7	29	12

MARCHES

TABLE II

PERCENTAGE DISTRIBUTION OF GRADES ACHIEVED IN ALL MATHEMATICS
COURSES ENROLLED FOR EACH GIVEN RANGE OF PERCENTILE SCORES
ON THE MARYLAND ALGEBRA APTITUDE TEST

Percentile scores	Total per cent	Percentages				
		A	B	C	D	F
80-100	100	15	29	37	11	8
60-79	100	6	19	39	26	10
40-59	100	4	13	43	23	17
20-39	100	6	6	6	63	19
0-19	100	10	20	35	20	15

TABLE III

TABLE III

PERCENTAGE DISTRIBUTION OF GRADES ACHIEVED IN ALL MATHEMATICS
COURSES ENROLLED FOR EACH GIVEN RANGE OF PERCENTILE SCORES
ON THE READING COMPREHENSION TEST

Percentile scores	Total per cent	Percentages				
		A	B	C	D	F
80-100	100	16	23	35	20	6
60-79	100	15	18	41	13	13
40-59	100	4	27	27	20	22
20-39	100	14	12	33	33	8
0-19	100	7	17	37	26	13

FIGURE 1.

A Comparison of the Observed and Theoretical Relationships Between Entrance Test Percentile Scores and Achieved Mathematics Course Grades

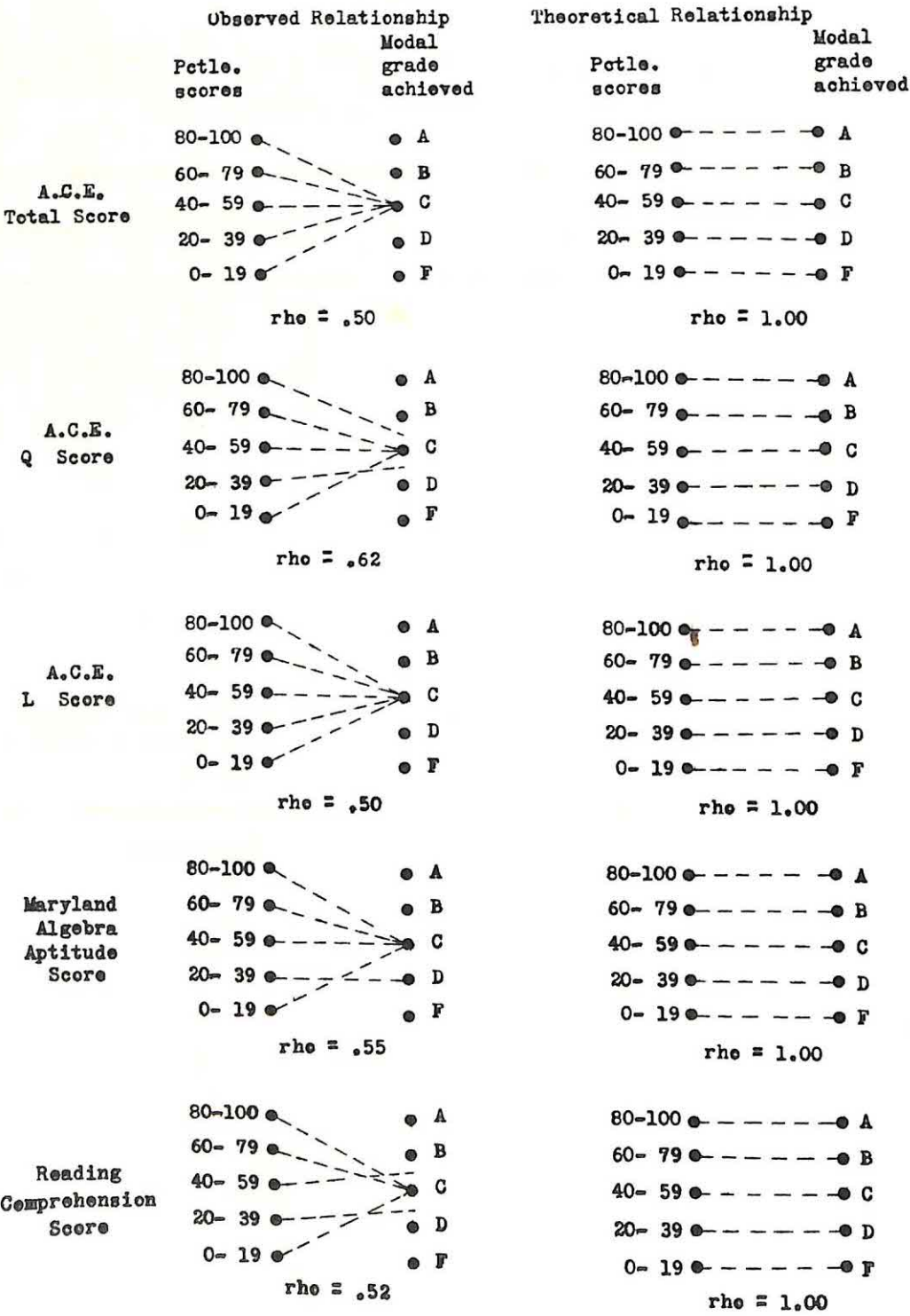


TABLE IV
THE PERFORMANCE PROFILE

Math. course title	Level of success typified	Name of Test and Percentile Score Grouping																				Level of success typified	Math. course title
		A.C.E. Test Component						Md. Algebra						Reading									
		Total Score			Q Score			L Score			Aptitude			Comprehension									
		Percentile Group			Percentile Group			Percentile Group			Percentile Group			Percentile Group									
		0-19	20-39	40-59	60-79	80-99	100	0-19	20-39	40-59	60-79	80-99	100	0-19	20-39	40-59	60-79	80-99	100				
All Math. courses	High Moderate Failure	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	High Moderate Failure	All Math. courses	
Introd. Algebra	High Moderate Failure	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	High Moderate Failure	Introd. Algebra	
Solid Geometry	High Moderate Failure	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	High Moderate Failure	Solid Geometry	
Algebra	High Moderate Failure	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	High Moderate Failure	Algebra	
Trig. and Analytic Geometry	High Moderate Failure	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	High Moderate Failure	Trig. and Analytic Geometry	
Plane Trig.	High Moderate Failure	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	High Moderate Failure	Plane Trig.	
College Algebra	High Moderate Failure	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	High Moderate Failure	College Algebra	
Analytic Geometry	High Moderate Failure	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	High Moderate Failure	Analytic Geometry	
Calculus	High Moderate Failure	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	High Moderate Failure	Calculus	

* Levels of success in Mathematics typified by College of Education Majors classified by entrance test. Score Ranges (N = 100 cases who entered the University of Maryland during the years 1950 to 1955).

TABLE V

A CHI-SQUARE TEST BETWEEN DICHOTOMIES OF SUCCESS OR FAILURE IN
GENERAL MATHEMATICS PERFORMANCE AND HIGH OR LOW
ENTRANCE TEST SCORES*

Entrance Test Name	Score Classification	Mathematics performance classification Frequencies		Chi-square value	Probability level
		Success	Failure		
A. C. E. (Total)	Upper-half	109	13	3.154	.08
	Lower-half	66	16		
A. C. E. (Q)	Upper-half	106	13	2.758	.10
	Lower-half	72	17		
A. C. E. (L)	Upper-half	106	15	1.183	.22
	Lower-half	69	15		
Maryland Algebra	Upper-half	149	20	4.560	.04
	Lower-half	30	10		
Reading Compre- hension	Upper-half	85	14	.377	.55
	Lower-half	61	13		

* Success refers to a grade of "C" or better. Failure refers to a grade of "F". Upper and Lower Entrance Test score classification refers to percentile scores above and below P50, respectively. A grade of "D" was interpreted as constituting neither success nor failure and was not included in the tabulation.

education majors and their entrance test scores is provided by the performance profile. This shows the achieved level of success characteristic for percentile score groupings on each of the entrance tests. The observed pattern suggests that moderate performance or "C" typifies many of the high entrance test percentile score categories. The legend for the profile is as follows:

High Success typifies a group when either the mode, median, or both represent a grade of "A" or "B".

Moderate Success typifies a group when either the mode, median, or both represent a grade of "C".

Failure typifies a group when either the mode, median, or both represent a grade of "F". (The letter grade "D" was interpreted as constituting neither success nor failure and was not included in the tabulation.)

The Chi-Square Test (2). To further test the predictive qualities of the entrance tests considered in this study, the distributions of percentile scores and performance letter grades were dichotomized. This provided a broader base by which to observe predictive qualities. Upper-half and lower-half percentile scores were compared against success or failure in mathematics. Chi-square values and their respective probability level appear for each entrance test.

Summary Observations

1. Tables I, II, and III indicate that the percentage distribution of letter grades in mathematics is frequently inconsistent with percentile rank position of entrance test scores. High ranking scores are often accompanied by a "C" grade level of performance and frequently "failure". On the other hand, low ranking scores on these preliminary tests are not always accompanied by "failure".

2. Moderate success in mathematics can be expected for the lower ranges as well as the middle and upper ranges of percentile scores for all of the tests considered.

3. Predictions of high success in mathematics cannot be safely predicted for the 80-100 percentile score range. In 25 out of 45 cross-tabulations of the performance profile, representations of moderate to low performance were also found to be typical for this group.

4. Even when intervals of achievement are ex-

tended to form broad dichotomies such as success and failure in mathematics versus upper-half and lower-half percentile scores, only one of the five entrance test measurements shows promise of being a useful instrument in predicting success or failure in mathematics course work. In this test relationship, only the Maryland Algebra Aptitude test has a sufficiently high chi-square value that is statistically significant beyond the conventionally accepted .05 probability level.

5. The disparity between the observed and the theoretical achievement levels, the performance profile, and the chi-square test all suggest motivational and interest factors as intervening variables.

6. For accuracy in prediction using the A.C.E., the Maryland Algebra, and a reading comprehension test, it appears essential that percentile scores be interpreted in terms of empirically established levels of performance peculiar to subgroupings of students. It is hypothesized that subgroups based on age, sex, marital status, etc. are characterized by different expected levels of achievement for the same entrance test score rank.

FOOTNOTES

1. Percentile rank scores on the A. C. E., the Maryland Algebra, and a reading comprehension test were used in this study. Mathematics courses most frequently taken by education majors were considered.
2. No attempt is made here to discern between tests that are primarily of an aptitude nature and tests which are primarily measures of achievement or a combination of native ability and achievement measurement.

REFERENCES

1. Bingham, Walter Van Dyke, Aptitudes and Aptitude Testing. N. Y.: Harper and Brothers, 1937, pp. 60-69.
2. Edwards, Allen L., Experimental Design in Psychological Research. N. Y.: Rinehart and Co., Inc., 1950, pp. 85-87.
3. McNemar, Quinn, Psychological Statistics. N. Y.: John Wiley and Sons, Inc., 1949, pp. 97-98.
4. Mowrer, O. Hobart, Learning Theory and Personality Dynamics. N. Y.: The Ronald Press Co., 1950, p. 198.

AN INVESTIGATION OF AN EXPERIMENTAL FIRST GRADE PHONICS PROGRAM

RICHARD H. BLOOMER
State University Teachers College, Geneseo, New York

CONSIDERABLE literature is available pertaining to various methods of teaching reading. One of the major areas of controversy has been the area of phonics. Questions of when to teach phonics, and of what kind of phonics to teach are of prime importance. Gates and Russell (5) and Tate, Herbert, and Zeman (6) find that formal phonics drill tends to be less effective than natural or intrinsic type phonics. Agnew (1) suggests that while the number of errors made by children who learn by phonetic methods is smaller, these children tend to read more slowly than others. The common conclusion from such data is that children should learn to read by means of developing a sight vocabulary at the outset; they should later learn some intrinsic phonics.

While a six and a half years mental age is generally recognized (3) as the age at which most children are ready to read by a phonic method, Gates (4), indicated that this may be a function of the teaching method. The present study represents an experimental test of this position.

The Experimental Method

The experimental method for the present experiment was designed to keep the advantages of early and concentrated phonics training, namely, independence in reading, few errors, and better spelling, and at the same time, the method attempted to mitigate the problems of lack of motivation resulting from drill and slow speed and lack of comprehension. Essentially the experimental method began as a modified formal phonic method, and shifted to a sight vocabulary method of teaching reading.

The experimental phonics method was designed in such a way as to restrict the words which the children learned to those which could be worked out phonetically. This notion of phonetic consistency followed a suggestion by Bloomfield (2). Each letter and the meaningful permutations and combinations of that letter with each preceding letter was presented systematically to the children. The sequence of presentation of the letters was such that succeeding letters were very dissimilar from each other in both sound and shape. Words were chosen

for the children to read on the basis of length and phonic complexity rather than frequency of occurrence. This technique served to expand the children's meanings beyond usual concepts. Work book material was presented which emphasized the building of meanings for these new words, and the children were taught the process of figuring out words from their sound components.

A common criticism of the letter-by-letter phonics method is that the children tend to become bored by the kinds of drill and exercise. The present study was designed with this limitation in mind, and the exercises were varied systematically and the limitation of drill to meaningful situations mitigated this problem. The procedure was instituted following a reading readiness program from a basal reader series, and lasted sixteen weeks. Subsequent to the phonic training, the children began their regular work in a basal reader series which continued for eight weeks until the final testing. The children were not taught the phonics in the basal reader program, but used the phonics they had learned with the experimental method.

The Control Method

The control teacher followed a regular basal reading program for the entire year. Following the readiness program the children learned between fifty and sixty sight words and then began to work in three primers. The words were learned primarily by the flash-card presentation method, with the recommended variations of the basal reader series. The children continued in primers and in first readers and were taught divided into three groups according to their reading skill. As the reading progressed, the children were given work in intrinsic phonics as prescribed by the basal reader series.

The teachers of each group spent approximately sixty minutes a day at various times in actual reading teaching. The other work of the children was essentially the same, and followed prescribed commercial programs.

Each of the teachers in the experiment possessed a bachelor's degree and had done some graduate work. The experiment represented the first year

of teaching for the experimental teacher. The teacher who taught from the basal reader series had taught for four years.

The children of an upstate New York consolidated school were divided randomly into first grade classes. During the first eleven weeks of the experiment the children went to school in a double shift, afternoons only. For the last five weeks of the experimental program and for the succeeding eight weeks, the children were on full-day schedules.

The children were pre-tested by means of a Gates Reading Readiness Test at the end of their reading readiness exercises. No significant differences were found between the classes. (See Table I) Two final tests were given. The Gates Primary Reading Test, Form 1, Type 1, Word Recognition was given at the end of the experimental procedure. The children were also given a spelling test at the end of the experimental sessions. Eight weeks subsequent to the previous testing, the children were given the Gates Primary Reading Test, Form 2, Types 1, 2, 3. Each of the teachers was aware of the experiment and was aware that measurement of her class would be taken. The control teacher was not aware of the experimental method.

Results

The initial testing. The results are presented in two forms: first, the percent of children achieving on given grade levels in the two groups; second, the reading test scores.

Table II indicates that no child in the experimental group failed to achieve at least second grade level in word recognition by the end of the experimental sessions. Fifteen percent of the experimental group were reading at grade three or better. In the control group thirty-four percent of the children were reading at that time on first grade level. A larger percentage of the control group was reading on third grade level or better. This indicates that the variation in the control group was at this point considerably larger than the range in the experimental group. Table III indicates significant F ratios.

The experimental group spelled more words correctly on a twenty-five word spelling test which was taken from words in the basal spelling program and from the reader series used for all groups. This difference is not significant; however, the experimental group had no spelling training during the experimental sessions except as a part of the reading program.

Final testing. Both the experimental group and control group averaged well above second grade level at the end of the experimental year. Table IV indicates the percent of children achieving at various grade levels on the final test. The experimental group had a smaller number of children reading at first grade level for all sub-sections of the test and for the total. In addition a larger number of

the experimental class was reading above third grade level for each section. Table V indicates that these differences were significant in favor of the experimental group in the Word Recognition and Sentence Reading sub-tests. The mean differences for paragraph reading and average scores were not significant.

F ratios for raw scores indicate that the dispersion of reading abilities for the experimental group are significantly smaller than for the control group. When grade equivalences are considered, the F ratios are significant except for the paragraph-reading sub-test. The data indicates that the experimental method produced better results in word recognition and in sentence reading than does the control method. Further, it indicates that the skills the children in the experimental classes tend to be more uniform.

Discussion

In addition to producing significantly better test results the experimental method added an increment in terms of efficiency of teaching to the reading program. In the experimental group in the eight weeks subsequent to the phonic training, each of the children in the experimental class read the pre-primers, primers, and first readers of the basal reader series. In the control group, who had been working on the same material for sixteen additional weeks, seventeen percent of the children did not finish the first reader. Likewise, in spelling, the experimental group completed their spelling work books for the first year in two weeks. The control group spent twelve weeks covering the same material. These differences appear to be too large to be simply a function of teachers especially considering the relative lack of classroom experience of the experimental teacher.

In the eight weeks between the first testing and the final testing, the experimental group gained in word recognition considerably more than did the control group. Two factors may be responsible for this marked change. First, the experimental method may have made the children more receptive to learning during the final part of the school year; or second, the three students who were not present at the initial testing may have been exceptional. An examination of the average word recognition scores for these three students on the final test indicates that the second is not the case. The average for these three students for the Type 1 Section of the test (Word Recognition) is 29.0 answers correct. This score is well below the mean on the final testing. It would suggest that the mean difference at the time of the initial testing, had these children been included, would have been smaller, but the gains in the final months of the school year by the same token, can be attributed to the experimental method. The data indicates that homogeneity of classes may in some measure be a function of method. The children in the experimental classes were taught as

TABLE I
READING READINESS SCORES

	\bar{X}	S^2	F	T
Experiment	21.00	23.54	1.04	.88
Control	19.83	24.48		

TABLE II
END OF EXPERIMENT TESTING. PERCENT OF CHILDREN
ACHIEVING AT VARIOUS GRADE LEVELS

	Word Recognition		
	1	2	3
Experimental	0	85	15
Control	34	42	24

TABLE III
 END OF EXPERIMENT TESTING. MEANS, VARIANCES, AND F RATIOS FOR
 SCORES AND GRADE EQUIVALENTS ON GATES PRIMARY READING
 TESTS, FORM 1, TYPE 1, AND SPELLING TEST

	Experimental (N = 29)	Control (N = 29)
Word Recognition, Raw Scores		
\bar{X}	28.38	24.00
S^2	46.96	157.48
F		3.35**
T		1.71
Word Recognition, Grade Level Equivalents		
\bar{X}	2.53	2.31
S^2	.11	.35
F		2.15*
T		1.65
Spelling Scores		
\bar{X}	13.19	10.31
S^2	44.76	75.57
F		1.62
T		1.41

* F .05 = 1.90

** F .05 = 2.49

TABLE IV
PERCENTAGES OF CHILDREN WHO ACHIEVED AT VARIOUS GRADE LEVELS
ON THE FINAL GATES PRIMARY READING TESTS

Grade Level	Group	N	Word Recognition	Sentence Reading	Paragraph Reading	Average Reading
1.0 - 1.99	Exp	32	0	6	12	0
	Cont	30	17	33	17	17
2.0 - 2.99	Exp	32	53	63	50	63
	Cont	30	53	46	60	56
3.00 - 3.99	Exp	32	47	31	38	38
	Cont	30	30	23	23	27

TABLE V
FINAL TESTING. MEANS, VARIANCES AND F RATIOS FOR
SCORES ON GATES READING TESTS

		N	Raw Scores				Grade Equivalents			
			\bar{X}	S ²	F	T	\bar{X}	S ²	F	T
Word Recognition	Experimental	32	34.97	66.43	2.71**	2.17 f	2.82	.73	2.20*	2.31 f
	Control	30	28.80	179.90			2.51	.382		
Sentence Reading	Experimental	32	34.03	63.29	2.13*	2.43 f	2.69	.234	1.99*	2.04 f
	Control	30	27.93	135.25			2.38	.466		
Paragraph Reading	Experimental	32	16.90	22.45	1.90*	.90	2.71	.317	1.46	.56
	Control	30	17.60	42.66			2.62	.465		
Average	Experimental	32					2.75	.208	1.88*	1.77
	Control	30					2.50	.392		

* F. 05 w/w 32/30 d.f. = 1.82

** F. 01 w/w 32/30 d/f/ = 2.34 f T. 05 w/w/ 32 d.f. = 2.04

a single group. The teacher made sure that each child was skilled with a particular letter before moving on to the next letter. The students tended to group themselves at the end of the experimental sessions, and at the end of the year there were four reading groups in the experimental classes. The differences between these reading groups, however, were much smaller than those in the control classes.

The reduction in range of reading ability within the experimental classes did not seem to hamper the faster students. This limited range results from the fact that none of the pupils in the experimental class were below the grade level.

Conclusion

The experimental method seems to substantiate Gates' (4) point of view that readiness for phonics at the age at which children derive benefit from phonics training is a function of method. A systematic reorganization of formal phonics training given prior to usual reading teaching techniques, then, appears to produce the superior results.

REFERENCES

1. Agnew, Donald C., The Effect of Varied Amounts of Phonetic Training On Primary Reading. Duke University Research Studies in Education, No. 5. Durham, N.C. : Duke University Press, 1939. 50 pp.
2. Bloomfield, Leonard. "Linguistics and Reading," Elementary English Review, 19: 125-30; 183-86.
3. Dolch, Edward William, and Bloomster, Maurine, "Phonic Readiness," Elementary School Journal 38: 201-5.
4. Gates, Arthur I., "The Necessary Mental Age for Beginning Reading," Elementary School Journal, 37: 497-508.
5. Gates, Arthur I., and Russell, David H., "Types of Materials, Vocabulary Burden, Word Analysis, and Other Factors in Beginning Reading, I and II," Elementary School Journal, 39: 27-35; 119-28.
6. Tate, Harry L., Herbert, Theresa M., and Zeman, Josephine K. "Nonphonic Primary Reading," Elementary School Journal, 40: 529-37.

THE RELATIONSHIP OF COLORS TO VARIOUS CONCEPTS

HARRY S. BECK*
University of Virginia
Charlottesville, Virginia

Introduction

In 1951 the writer conceived the idea of having subjects associate colors with various stimulus words as a means of evaluating personality and differentiating various diagnostic groups. The Word-Color Association Test was originally conceived as a projective technique for clinical use, but it soon became apparent that it might be adaptable for group use with very little loss of information.

The present study is the first in a series intended to validate the technique. The following hypotheses were selected for the initial investigation: (1) Certain colors are more commonly associated with a given concept than other colors, (2) certain colors are rarely associated with certain concepts, and (3) these associations will remain relatively stable over a period of time.

Rationale of the Test

(The rationale for the test is that color has meaning for the individual and that the affective value of a given color will remain constant at least for the duration of the testing period, which is less than one hour.) (It is further hypothesized that the color will serve as a common denominator linking the words associated to a given color together.) Thus, when an individual associates the color black to stimulus words such as: hate, authority, father, boss, and teacher, we may be able to infer that he dislikes authority and possibly may be in conflict with individuals in positions of authority.

This method has interesting possibilities as it provides a certain subtlety or ambiguity making it much more difficult for the subject to fake his responses. Using this technique, the subject is less likely to know what is being sought than he would if he were relating concepts directly to each other e. g., hate-father.

Other writers (1, 2, 3, 4, and 5) have presented evidence which suggests that such a technique is possible and give reason to believe that color has value for the individual biologically, emotionally, and socially or culturally.

(If then, specific colors can be related to specific concepts, this might be valuable to researchers investigating: Personality; the construction of tests

to evaluate attitudes towards various situations or interpersonal relationships, areas of conflict, tendencies toward depression, etc.; advertising and product packaging; home, school, office, and factory decorating; and communications or signaling to mention a few. It might also be useful in distinguishing various clinical groups such as psychotics, organics, delinquents, etc., from one another.)

The Subjects

The subjects utilized for checking hypotheses one and two consisted of 455 boys and 409 girls ranging from 12 to 16 years of age. They were all enrolled in public schools located in both urban and semi-rural communities. Although there were some negroes in the original group, they were not included in the 864 reported on here. In analyzing the data it became apparent that different norms would probably be necessary for whites and negroes.

The second group, used to check hypothesis three, was composed of 15 boys and 22 girls enrolled in a seventh grade of public school. This group was entirely separate from the subjects used in group one. Group two was also all white and within the 12 to 16 year limits of group one.

Although unproven objectively, it is assumed that the groups, especially group one, were normally distributed with respect to intelligence, achievement, social adjustment, etc.

Procedure

The word-color association tests were administered by the teachers to their classes as groups, with instructions to keep the groups under 35, if possible. In group one each child was given an answer sheet which contained identifying information and the names of eleven colors (see Table I for names of colors and stimulus words). Immediately below these were sixty blank numbered spaces for the subject to write in the colors associated with the stimulus words. The following directions were read to the subjects by the teacher: "This is a test to determine what colors are ordinarily brought to mind by various words. I shall read a list of words at the rate of one in 10 seconds, and I want you to write down on your answer sheet the first color that comes

*Footnote found at end of article.

TABLE I

NAMES OF COLORS AND STIMULUS WORDS
EMPLOYED IN THE WORD-COLOR TEST

Red	Pink	Purple	Black
Yellow	Orange	Brown	White
Green	Blue	Gray	
<u>Stimulus Words</u>			
1. auto	21. sister	41. bedroom	
2. cigarette	22. bathroom	42. son	
3. baby	23. liquor	43. doctor	
4. love	24. divorce	44. persistent	
5. friendly	25. party	45. daughter	
6. house	26. dream	46. afraid	
7. girl	27. brother	47. marriage	
8. hate	28. minister	48. mother	
9. work	29. hair	49. wicked	
10. farm	30. teacher	50. kiss	
11. shoes	31. job	51. hand	
12. happy	32. sofa	52. naughty	
13. father	33. boss	53. husband	
14. play	34. strong	54. angry	
15. girl friend	35. wife	55. hunger	
16. dislike	36. sick	56. dance	
17. boy friend	37. pleasure	57. family	
18. escape	38. unhappy	58. sex	
19. church	39. money	59. color liked best	
20. jealous	40. weak	60. color disliked most	

to your mind in connection with that word. Choose the color from among the eleven listed at the top of your answer sheet, and write the name of the color in the numbered space corresponding to the number of the word I read. Remember to indicate the very first color that you think of. Please write clearly and legibly."

The teachers were instructed to spell out the words son and weak to avoid confusion with sun and week. They were also instructed not to repeat words after the 10-second time limit expired, and not to go back if some had omissions. The subjects were not allowed to see the stimulus word list before or after the testing.

Group two was given the same type of test except that only twenty stimulus words were utilized. This group was tested twice, with the second presentation occurring one week after the first. The subjects were given no indication that a second test would be given in order to promote as much forgetting as possible, or at least avoid any conscious attempt to remember associations. At the time of the second testing, they were told that they were being given a similar test but that many of the words were different. Actually the two tests were identical. Although memory may have been a factor, it was felt that the lapse of one week's time plus the great number of possible combinations would minimize the memory factor.

The data from both groups were tabulated and converted into percentages by sexes to indicate the associations of given colors with stimulus words. The omissions were also tabulated. Chi squares were computed for each stimulus word to determine whether the color choices varied significantly from a chance distribution. In group one a binomial expansion was utilized to determine the probability of any given color being associated with any given stimulus word beyond chance expectation at the .01 level of confidence. In group two Chi square was also used to compare the results of the first and second testings.

Results

In both groups it was found that certain color choices varied from the expected frequencies at a level well beyond the .001 level of confidence, indicating that certain concepts (at least those used here) are significantly associated with certain given colors.

In group one, assuming that the subjects could either choose or not choose a given color, it was found that any color associated with a given stimulus word by 11.6% or more of the population exceeded chance expectations at the .01 level. If a color is associated with a stimulus word by three per cent or less of the population, it is considered a "rare" response.

Of the sixty stimulus words, 39 are associated with a given color by 25-49% of boys and 33 with girls. At the 50-74% level seven words are associated by boys and girls, and at the 75+% level only one

word is associated by boys and three by girls. By grouping the responses into various percentage levels, it is hoped that an objective numerical scoring system can be set up. It is also interesting to note that only five stimulus words were not associated with some color by more than three per cent of both boys and girls, which offers another possibility for scoring.

In group two the responses were totaled for each testing session and compared using the Chi square technique, with no significant difference being obtained. This would indicate that color associations remain stable over a one week period and thus would remain constant during the test session which consumes about 30 minutes time.

Conclusions

With regard to white public school children ranging from 12 to 16 years of age, our data indicate that:

1. Certain colors are associated with given stimulus words more often than could be expected on a chance basis.
2. Certain colors are rarely (3% or less of subjects) associated with certain stimulus words.
3. Word-color associations tend to remain stable over a period of at least one week.
4. The word-color association technique shows possibilities for objective scores for group use.
5. Sex differences between color preferences associated with given stimulus words were demonstrated.
6. Failure to associate some color with the stimulus word rarely occurs with most stimulus words.

FOOTNOTE

- * The writer is greatly indebted to Dr. Goldine C. Gleser, University of Cincinnati Medical School for her time, interest, and suggestions regarding the analysis of the data.

REFERENCES

1. Werner Wolf. The Expression of Personality. New York: Harper, 1943, pp. 277 ff.
2. Louis Cheskin. Colors - What They Can Do for You. New York: Liveright, 1948, p. 35.
3. Josefina Rodriguez. "Las Preferencias de Color en los Niños de Edad Escolar," Revista de Psicología General y Aplicada. Madrid: 1950, 5, 59-67.
4. Lawrence E. Abt and Leopold Bellak. Projective Psychology. New York: Alfred A. Knopf, 1950, p. 442.
5. T. Obonai and T. Matsuoka. Manual of the Color Symbolism Personality Test. Nakayama, Shoten, Tokyo, 1956.

RESEARCH INSIGHTS INTO THE RELATIONSHIP BETWEEN TEACHERS' ACCEPTANCE ATTITUDES, THEIR ROLE CONCEPTS, AND STUDENTS' ACCEPTANCE ATTITUDES

ROBERT E. CUMMINS

Arkansas Agricultural and Mechanical College

IN THEIR FERVENT insistence that American youth learn some "facts" thoroughly, many school patrons tend to ignore or minimize other aspects of growth and development, such, for example, as emotional and social growth. It can easily be inferred from the current criticism of schools that the most important--perhaps the only--competence required of the teacher today is that she be "well-rounded." This is interpreted to mean that she must know her subject matter thoroughly and have broad intellectual interests. Whether she likes herself or her children are factors seldom mentioned. School administrators themselves, desperate for personnel, rely primarily on college course grades, length of experience, and casual interviews in determining competence.

The emotional adjustment and social attitudes of the teacher receive attention only when they eventuate in extreme--even violent--misbehavior. This obtains in spite of the fact that there is considerable evidence to indicate that the emotional climate of the classroom affects not only the learning of "facts" but also, of greater significance, the emotional health of a student. And the teacher is the prime determiner of the emotional climate of the classroom.

It is the purpose of this article to report two findings revealed by a study of the relationship between administrative climate and teacher capacity for professional growth.¹

The study involved four principals, 92 high school teachers and 119 12th grade students. The schools were 1-12 grades, in communities of 2,000 people. The teachers were matched as to age and experience factors.

The first hypothesis relevant to this report predicted a significant relationship between the teacher's acceptance of self and others and the student's acceptance of self and others.

Data from School "A" included 25 teachers and 48 seniors; School "B", 22 teachers and 71 seniors.

The Index of Adjustment and Values² was used to probe self-other attitudes. The data revealed that the students of School "A" differed significantly from the students of School "B" in their self-other attitudes. The teachers of School "A" differed significantly from the teachers of School "B" in their self-other attitudes. The conclusion seems justified that there does exist significant relationship between teacher acceptance of self and others and student acceptance of self and others.

The second relevant hypothesis predicted a statistically significant relationship between a teacher's acceptance of self and others and her perception of her role as a teacher. Ninety-two teachers provided the data for this aspect of the study. The Index of Adjustment and Values and a teacher role concept Q-sort developed on a continuum from accepting--permissive to rejecting--introjecting teaching behaviors were employed to gather the data.

The discrepancy between "self" and "ideal" as revealed by the Index yielded the self-others acceptance score of these teachers. The Pearson product of coefficient correlation between the role perception of the teacher and the composite opinion of a "jury" of six college professors as to the teachers' role served to provide a statistical measure of their perception. The results were statistically significant.

Of the ninety-two teachers, the Index identified seven as persons who rejected both self and others. An analysis of role sorts of these seven revealed that each had delineated a role characterized by definite misanthropic attitudes and behaviors.

These seven teachers placed the following items in the upper end of the continuum claiming them as most characteristic of their teaching.

- Rejects those students who do not like her.
- Puts student to the test whenever possible to strengthen him.
- Spurs student to greater effort by making him ashamed of his inadequacies.

* Footnotes will be found at the end of this article.

--Introduces the element of competition into her classes.

--Anticipates student efforts to cheat on examinations.

--Conditions students to face the hard realities of adult life.

--Protects students from a natural tendency toward delinquency.

--Punishes student in proportion to the seriousness of his offenses.

--Keeps her desk and teaching materials securely locked while not using them.

In the aggregate, these items delineate personalities possessed of dangerous potential for blighting the lives of the young people entrusted to them in the intimacy of the instructional process.

The fact that the study betrayed "only" seven out of 92 such people should occasion little complacency. Even one misanthrope is too many in light of the fact that perhaps as many as 175 young people are exposed to this pernicious influence seven hours

of a day, 180 days a year.

In the interest of a clearer understanding of these findings, it should be pointed out that this effort was in the nature of a pilot study. No claim is made to a cause-effect relationship between these factors. It is intended merely to point out that here appears to be a fruitful area of investigation capable of revealing highly meaningful insights into the critical teacher-pupil instructional relationship.

FOOTNOTES

1. Robert E. Cummins, "An Evaluative Study of Certain Teacher Perceptions Related to Professional Growth", an unpublished doctoral dissertation, Alabama Polytechnic Institute, Auburn, Alabama, August, 1957.
2. Robert E. Bills, "A Manual for the Index of Adjustment and Values", Alabama Polytechnic Institute, Auburn, Alabama, 1958.

CHANGES IN ETHNIC REACTION TENDENCIES DURING HIGH SCHOOL

JOHN D. McNEIL
University of California, Los Angeles

THE EDUCATIONAL program of most schools aims at the improvement of inter-group relations and the reduction of negative stereotypes or prejudice toward others. Unfortunately, not too much is known about the effectiveness of schools in achieving these expected outcomes. It is intended to describe here one school's approach to such an assessment, revealing changes in ethnic attitudes during high school.

The literature on prejudice and ethnic relations indicates that while a large part of the information presented in the classroom becomes part of the ethnic beliefs of students, favorable changes in attitudes are likely to ensue only when the following conditions are present: 1) the information presented is more favorable to the group in question than the previous belief of the average student, 2) the instructor indicates directly or indirectly that his own attitudes toward the group are more favorable than their own previous attitude, and 3) the relationship between student and instructor is such that students tend to accept the instructor's feeling and action orientation toward the group in question.¹ Education may or may not have a liberalizing effect upon the stereotypes of students, depending on the personality structure and social situations of the individual concerned. Displaced hostility is presumed to be associated with frustration and insecurity. Although ethnic attitudes remain subject to modification by a variety of social influences, some believe that by the age of fourteen these attitudes are crystallized in the vast majority of children.² Interestingly, the most extensive study of prejudice during the high school years reported a steady increase in prejudice from the seventh through the twelfth grade.³

The present study had two phases: 1) the comparison of number and dimensions of prejudiced stereotypes expressed by fifty entering tenth grade students and fifty graduating twelfth grade students in the same institution and 2) the comparison of stereotypes expressed by students when they were at the tenth grade level with the attitudinal responses of the same students when at the twelfth grade

level.

Procedure and Results from First Phase

Pursuant to the first phase, students representing twenty-five percent of those in the respective grade levels were selected. The sampling provided for a distribution of ability as indicated by composite Iowa test scores of general educational achievement. The instrument for eliciting expressions of prejudice consisted of a list of forty words, such as lonely, artists, Mexicans, etc. Each item was prefaced by the word "most." Respondents were asked to complete a sentence from each item, expressing their real feelings and beliefs. Eight of the forty items appeared as relief from more stimulating words.

Responses were independently read and categorized by three judges who were guarded against permitting their own cultural frame of reference to bias the analysis. Only those reaction tendencies which were predominately unfavorable were sought. These negative responses were placed in the categories of prejudice suggested by Kramer.⁴ The latter system recognizes the cognitive (unfavorable belief), affective (unfriendly feeling), and conative (desire to see situation of others worsened) dimensions of prejudice. Within this framework, responses were further identified. For example, the expression "Most Scotchmen are money-grabbing" was placed within the specific category "acquisitive."

Results of the first phase indicated that twelfth graders were more negative in their reactions than tenth. Stereotypes against racial (Negroes, Mexican, Chinese) and religious groups (Roman Catholics and Methodists) were more than double those of tenth. Teen-age intelligentsia ("Brains") and teen-age non-conformists ("Kats") were especially disliked. The upper ability members of the class expressed more unfavorable stereotypes toward such occupational groups as artists, sailors, labor leaders, and the like than did others. Categories of total tenth and twelfth grade responses were most

* Footnotes will appear at the end of the article.

frequently expressed in this order:

cognitive dimension--seclusive, acquisitive, impulse gratifying, inferior intellectually, inferior status, threatening, impulse non-gratifying;

affective dimension--non-sympathy, distrust, disgust, fear;

conative dimension--withdrawal, attack.

Seventeen percent of all responses were negative.

Procedure and Results from Second Phase

The second phase of the study required administering two and one half years later the same instrument to the same students under like conditions. Twenty-eight from among the original group of fifty tenth grade students were available for this purpose.

The findings from this second inquiry were very similar to the first. As twelfth graders, students tended to express more prejudices. Again, seventeen percent of all responses were negative. Further, students tended to retain their relative positions to each other with respect to the number of negative stereotypes expressed. Those who were most prejudiced in the tenth grade were likely to be most prejudiced in the twelfth. (This correlation was significant at the .01 level.) Stable and consistent patterns of stereotypes toward particular items tended to be maintained.

The coefficient of correlation of tenth and twelfth grade responses associated with both items and categories was .79 and .83 respectively (significant beyond the .01 level). This pervasive nature of responses did not hold for all items. Unusual increase in prejudice against labor leaders was indicated by a rank order change for this item from eleventh to second position during the two and one half year interval. It might be presumed that unusual economic and political relationships surrounding labor leadership at this time influenced the stability of prejudice expressed by students.

Usefulness of the Instrument

Comment upon the usefulness of the "Most Test" as an instrument for identifying prejudice is in order. It has the advantage of not crystallizing in the mind of the student heretofore unformulated answers. It possibly does not reveal conative dimensions of prejudice as well as do social situation tests. However, since conative components usually require a policy orientation, teen-agers may express fewer conative prejudices due to their lack of experience in making policy. The reliability of the instrument suggests its value as a predictive device.

While this study revealed a general increase in the amount of prejudice during high school, one cannot say that the schools are failing. The fact that a few individuals showed lessened prejudice is encouraging. For others, prejudice might have increased to a greater extent had it not been for the school's efforts. The study points to the need for 1) distinguishing between the effects of age and education, 2) seeking the relationship between expressed paper and pencil reactions and performance in inter-cultural relations, and 3) indicating teaching arrangements most appropriate for eradicating prejudice among particular individuals.

FOOTNOTES

1. Harding, John, et al., "Prejudice and Ethnic Relations," Handbook of Social Psychology, Gardner Lindzey, editor, (Cambridge, Mass. Addison-Wesley Publishing Company, Inc. 1954) p. 1036.
2. Ibid., p. 1036.
3. Minard, R. D., Race Attitudes of Iowa Children, University of Iowa, 1931, 4, No. 2.
4. Kramer, B. M., "Dimensions of Prejudice," Journal of Psychology, 1949, 27, 389-451.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 53

FEBRUARY 1960

NUMBER 6

CONTENTS

Summary of Investigations Relating to Reading July 1, 1958 to June 30, 1959—William S. Gray.....	203
Measurement of College Level Reading Competence in a Content Area—Irma T. Halfter and Frances M. Douglass.....	223
Changes in Affect Attributable to Instruction in Reading Improvement at the College Level—Meryl E. Englander.....	231
Survey of the Research Related to the Reading Ability of the Gifted—Sister Josephina, C.J.S.	237



DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

SUMMARY OF INVESTIGATIONS RELATING TO READING JULY 1, 1958 TO JUNE 30, 1959

WILLIAM S. GRAY
University of Chicago

AN UNUSUALLY large number of studies of reading were published during the year. In the past a generous policy has been followed in listing school or classroom studies which did not meet all requirements of scientific research. This was done in the belief that every encouragement should be given to those who seek to find solutions on an objective basis to reading problems. The number of such studies is increasing so rapidly that it is no longer possible to include reference to all of them in this annual summary. Those referred to this year represent either a new or distinctive approach to a problem or suggest significant issues in need of further study.

Fay (27) brought together a series of interesting classroom studies of the type referred to above. They were carried on by teachers who took an extension course offered by the University of Indiana and supply concrete evidence of the value of research by classroom teachers. Following the pattern set by Fay an annual summary of such studies on a nation-wide basis would be very helpful, particularly if followed by constructive suggestions concerning techniques of research to insure more valid results.

The two most noteworthy reports published during the year were prepared abroad. One was entitled Silent Speech and Silent Reading. It was prepared by Dr. Ake W. Edfeldt (22) of the Institute of Educational Research, University of Stockholm, and will be discussed later. The second was entitled Backwardness in Reading: A Study of Its Nature and Origin. It was prepared by Dr. M. D. Vernon, (108) University of Reading, England, and summarizes the results of numerous pertinent studies under the following headings: "visual perception in reading"; "auditory perception in reading"; "the nature of reading disability"; "innate factors as causes of disability in reading"; "the relation of various acquired defects to reading disability", and "the effect of environmental factors on reading ability and disability".

Summaries Relating to Specific Aspects of Reading

The number of scientific studies relating to reading is now so extensive that those interested in spe-

cific problems find it desirable to summarize from time to time the results of all related studies. Summaries of this type published during the year are listed below and should be studied in detail by those interested.

"Critical Reading in the Content Area," by Artley (2).

"Visual Handicaps to Reading," by Eames (20).

"Persuasion and Personality: Readers' Predispositions as a Factor in Critical Reading," by Eller and Dykstra (23).

"New Approaches to the Study of Interpretation in Reading," by Gray (41).

"Summary of Reading Investigations, July 1, 1957 to June 30, 1958," by Gray (42).

"Some Recent Research in Visual Problems in Reading," Huelsman (50), organized under three headings: "outline form perception", "use of the tachistoscope", and "general visual achievement".

"Some Research on the Impact of Reading," by Russell (86). The specific topics discussed are "characteristics of the reading material", "content of ideas", "characteristics of the reader", "overt responses as effects of reading".

"The Nature of Critical Reading," by Sochor (93).

Good Reading for Poor Readers, by Spache (94).

"Recent Studies of Eye-movements in Reading," by Tinker (103).

Backwardness in Reading: A Study of Its Nature and Origin, by Vernon (109).

"Studies in Listening, I: Relative Values of Oral and Visual Presentations," by Witty and Sizemore (117).

"Studies in Listening, II: Relative Values of Oral and Visual Presentations (Lectures, Movies, Examinations, and Advertising Materials)" by Witty and Sizemore (118).

"Studies in Listening, III: The Effectiveness of Visual and Auditory Presentations with Changes in Age and Grade Levels," by Witty and Sizemore (119).

The Sociology of Reading

The social role and implications of reading challenge the interest of a small group of researchers

to carry on pertinent studies.

Use of different mass media. A survey by McEvoy (70) of the impact of mass media on 260 homes of Indiana, largely middle-class urban, showed that 92% of the families contacted had at least one reader of newspapers; 71% had at least one radio listener; and 78% of the families were reached by television. Those interviewed preferred television if they could have access to only one medium. Interpretation of the data indicated that similar studies should be made for "the central city of each trade area that sends out mass media." The findings also indicated that the vitality of the content of newspapers is of great importance in securing readers.

A nation-wide survey by the Institute of Student Opinion (51) among 10,149 teen-age students from all parts of the country showed that about 78 per cent of them read newspapers, exclusive of comic pages, the preceeding day, devoting about 15 minutes, more or less, to them. About 45 per cent spent some time reading a non-required book, the modal amount of time devoted to it varying from one-half to an hour. About 80 per cent watched TV, varying in time from less than 30 minutes to 6 hours or more. The mode was from two to four hours. Whereas about the same percentage engaged in newspaper reading and television viewing, the amount of time devoted to the latter greatly exceeded that devoted to the former. It is significant that less than 25 per cent spent time reading comic books.

Effect of various forms of presenting news and of bias in interpretation. Publishers as well as distributors of news through other mass media are concerned with the most effective modes of presenting news. Mehling (73) analyzed the responses of seven groups of university students to determine the relative effectiveness in changing personal attitudes of a photo-caption combination and a story alone. A comparison of the results of pre-tests and post-tests showed that the former was the more effective. No greater susceptibility to influence on the part of one sex over the other was revealed.

Kerrick (54) secured responses from 50 college women "to explore the effect (on meaning) of combining pictures and captions", one conveying a positive influence and the other a negative. The findings showed that "if the picture and caption are similar in original meaning, these element-meanings are added together in the combination to produce a more extreme judgment." If they are not similar in original meaning, "these element-meanings compete for dominance."

Engel, O'Shea and Mendenhall (25) secured responses from 233 college students to test two hypotheses: (a) "People read and interpret statements dissimilarly Ego-involved individuals perceive reading matter differently from individuals without commitment in the area covered by the material"; (b) "The direction of interpretation of ambiguous reading matter is determined by the existing bias of the individual." An analysis of the responses

tended to support the hypotheses. The findings also suggested that the journalistic use of ambiguous statements "may appeal to the existent biases of the readers and lead, if anything, to the strengthening of prejudice." The investigators urged greater clarity in published materials as an aid "to a much greater meeting of the minds."

Trend in newspaper circulation. Peterson (78) compared the total circulation of daily newspapers with the total population of the United States and with that over 20 years of age for the period 1929 to 1957. The general fact revealed was that whereas there is a fair relationship between percentage increases in population and that in newspaper circulation, the latter failed definitely to equal the former. Likewise it did not keep pace with the percentage increase in households. The factor of income bore a positive relationship to circulation. Peterson questioned the adequacy of newspaper circulation today in view of the need for increased information and knowledge. He suggested two possible explanations for the steadily downward trend since 1950 in newspaper circulation per household, namely the advent of another mass media or reduced efforts by newspapers to sell their product.

Emphases in publications serving special groups. Two investigators were concerned with the interests emphasized by publications for special groups. Fishman and Fishman (30) made "a social-psychological analysis" of editorials in five foreign-language newspapers in New York City to determine the extent to which they dealt chiefly with their "own ethnic problems" or with those relating to "adjustment integration". He found that editorials in newspapers for recent immigrants give large emphasis to ethnic problems with increased emphasis on adjustment-integration the longer they are here. Garver (34) analyzed the content of 34 international trade union publications to determine their chief purpose. He concluded that "the labor publication today is published primarily to sell the union leaders and the organization to the worker". These findings indicate that what one reads is often designed to serve special interests and to cultivate distinctive attitudes.

Changes during a decade in non-fiction magazine articles. Ellison and Gosser (24) made an analysis of the content of 9 non-fiction magazines for three-month periods in 1947 and in 1957 to identify changes, if any, in subject matter, writing style and article length. An assumption underlying the study was that "shifts in the nature of published articles may be taken as some indication of popular mood". The analyses made showed no shift toward brevity in the articles published during the ten-year interval. They revealed a growing concern on the part of the reader "about his own orientation and adjustment". For example, articles on personal management, physical and spiritual, increased significantly. Narrative style ranked high as the favorite. The investigators concluded that the current state of magazine articles "is one of abounding vitality". They found

"no deterioration in quality of workmanship, integrity of reporting, or dignity of topic".

The Physiology and Psychology of Reading

Of the various studies listed in the bibliography, more relate to these aspects than to the sociology or the pedagogy of reading.

Relation between reading and intelligence. The relationship between reading and intelligence continues to intrigue researchers. Hage and Stroud (44) found correlations between the scores of 800 ninth graders on the Lorge-Thorndike Intelligence Tests and the Pressy Reading Rate and Comprehension Tests and the Iowa Tests of Basic Skills. "Reading comprehension and reading rate correlated significantly both with the verbal and non-verbal intelligence scores, but more highly with verbal scores." Through the use of partial and multiple correlations they found that "verbal scores are affected more than nonverbal scores by reading proficiency". Furthermore, "verbal intelligence scores give a somewhat better prediction of academic achievement (reading) than do nonverbal scores".

On the basis of data secured from 513 fifth-grade pupils Maney (71) also concluded that there is a significant correlation between scores on group tests of verbal intelligence and general reading. Furthermore, "proficiency in literal reading interpretation of science material may be predicted with a fair degree of accuracy from scores on group tests of verbal intelligence". Kittell (58) secured evidence from 83 third-grade pupils which supported previous findings to the effect "that bilinguals suffer from a language handicap when intelligence is measured by verbal tests". However, the bilinguals did not differ significantly from the unilingual group in reading ability.

Correlations among reading, spelling and arithmetic abilities. To secure further insight concerning the relationships "between the spelling and general reading abilities of children in grades three through eight", Morrison and Perry (74) gave the California achievement and Mental maturity tests to 1,007 pupils. The correlation for the combined grades was .79. The correlations varies from .85 for the third grade to .75 for the eighth grade. These findings indicate "a fairly conclusive" relationship, the higher correlations being in the lower grades.

In a closely related study Waldman and Triggs (109) gave the Diagnostic Reading Tests: Survey Section and the Lincoln Diagnostic Spelling Tests to pupils in grades four to seven inclusive. They found correlations of .63 and above between word recognition and spelling, but less significant correlations in the case of comprehension. The correlation between spelling skills and vocabulary remained about the same throughout the upper grades but was not as high as those between word attack skills and spelling. Auditory discrimination bore "a very marked relationship to success in spelling".

Reed and Schonfield (81) criticized the conclusions in a report of Dr. R. Lynn in which he examines the "disparity between the reading and arithmetic attainments of school children". He also concluded that there is "positive association of anxiety with better reading than arithmetic" and that possibly "some children do not learn to read because they are not anxious enough". Reed and Schonfield expressed skepticism concerning these conclusions on three grounds: confusion in Lynn's use of clinical terms; inaccuracies in quoting and interpreting the literature; and questionable experimental procedures and erroneous statistical computations. In a reply Lynn (67) defends the conclusions of his previous article and points out misinterpretations of his article by Reed and Schonfield.

Relative merits of listening and reading. King (56) compares the comprehension scores made by 225 boys and 250 girls in the primary grades on comparable passages read and heard. The data thus secured showed that the boys scored higher on the auditory than the visual tests in eight out of ten passages, only three of the differences being significant. The girls scored higher on the auditory tests on five of the passages, only one of the differences being significant. When the results on all the tests were considered "no significant difference (was found) between the responses of boys and girls whether the test was given orally or visually". The influence of intelligence was not clear cut.

Westover (112) gave tests to 198 college students to determine if they differed in ability to answer objective-type questions if read to them or if they read them themselves. He found no group differences on the two types of tests; also "that student's preferences for listening or reading tests have almost no relation to their performance", and "that comparisons of groups on the basis of similarity of means and standard deviations often mask important differences in traits among individuals".

Haberland (43) gave listening, reading and intelligence tests to 110 college freshmen in order to "investigate" their listening ability. He found that the correlation between the scores on listening and reading tests varied considerably with the tests used. He attributed this to the fact that listening ability, as measured by different tests, varies widely. Sex differences were not significant for the total group tested. In contrast to the findings of King, quoted above, there was a marked relationship between scores on listening and intelligence tests. Whereas the results of the three studies reviewed agree in certain respects, the differences suggest the need of further research through the use of more carefully selected tests and procedures.

Nature and development of reading as revealed through eye-movement records. Five studies of reading through the use of photographic records of eye movements in reading were reported. Taylor (100) presented averages for each grade from one to six and for the junior high school, senior high school and college for five components of reading:

"fixations per 100 words"; "regressions per 100 words"; "average span of recognition"; "average duration of fixation"; and "average rate of comprehension". The data showed that as children mature "their mechanical and over-all functional efficiency in reading tends to improve." Whereas growth is most rapid below Grade V, it continues at least to the college level. Of special significance were the wide variations in the achievement of individuals reported at each grade level.

Gilbert (39) reported a genetic study of the number of fixations, regressions and durations of fixations per 100 words for 22 subjects when in the second grade and again when college juniors or seniors and for seven subjects at the second, fourth and sixth grade levels and in college. His subjects scored higher at each grade level than was true in the case of Taylor's subjects, due possibly to differences in the difficulty of the materials read. Gilbert found that all second-grade pupils were still quite inaccurate in moving the eyes from the end of one line to the beginning of the next. They also read predominantly word by word making many extra fixations. Some subjects established this pattern of reading so firmly that it persisted to the college level. Long duration of fixations in the second grade also persisted. The investigator interpreted these findings to mean that there are factors in speed and span of perception which are as yet not clearly understood and justify further research.

Gilbert reported the results of three additional studies of visual perception. The first (37) showed that mature readers made more perceptual errors when reading with eye movements than in an artificially contrived reading situation that did not require saccadic movements. Both good and poor readers made fewer errors without saccadic movements when the exposures were 2/24 than with saccadic movements when exposures were 4/24 of a second. Furthermore, slow readers seemed to need "a longer interval of time to stabilize their fixations to a point of maximum efficiency than is necessary for the fast readers". In the second study (38) Gilbert studied the influence of varying the time of exposure of a visual stimulus before the reader encountered an interfering one. He found that students are not influenced to the same degree "by an increased restriction in the period of freedom from interfering stimuli", that "the poor readers are influenced more than good readers by interfering stimuli", and "that the narrower the span of perception, the easier it is to avoid the influence of extraneous visual material". In a third study (36) Gilbert studied "the influence of interfering stimuli in perception of meaningful material". The data secured indicated that "the average college reader will need fixation pauses which will average better than 4/24 of a second in duration if he is to use a span of perception of more than one word and avoid interference from the stimuli resulting from succeeding fixation pauses".

Mental imagery in reading. The value of mental

imagery in reading has long been recognized. Fennema (28) questioned fourth-grade pupils following the reading of a passage to learn more about the nature and extent of their imagery and its relation to intelligence and reading ability. Emphasis in the interview was placed on visual imagery. She found that those "who formed the fewest images were often the fastest readers in the group". Two possible explanations were offered: some fast readers understand less than poor readers; fast readers may read with less imagery of the type measured. Both views are supported by the results of earlier studies. The latter explanation finds added support in Fennema's findings to the effect that bright children have less "mental" imagery than children of lower intelligence. Furthermore, the correlation between reading ability and extent of mental imagery was negative for one group and zero for another. The findings reported are limited in value due to the use of the more or less informal methods of measurement used and the limited types of imagery studied.

Factors influencing memory of words and of paragraph meanings. Wertheimer and Gillis (11) secured responses from 501 subjects to three lists of twelve words each, involving six variables (frequency, objectivity of referent, number of syllables, number of letters, sequence in presentation of words, characteristics peculiar to a given word) to determine their influence on lapse of memory. They found that "words which retain their meaning longer are one syllable rather than two, are short rather than long, perhaps have an objective rather than an abstract referent, and occur early rather than late in a sequence".

Ausubel and Blake (3) measured the effect on retention of the content of a passage about Buddhism of reading a passage containing the same content and in addition a discussion of the similarities and differences between Buddhism and Christianity. The data secured supported the hypothesis that "the explicit comparison method of learning reduces proactive inhibition by enhancing the discriminability of the learning material from the subsuming material".

Silent speech during silent reading. The extent and function of inner speech during reading have long been subjects of research. Illuminating new evidence has been reported by Faaborg-Andersen and Edfeldt (26) who secured electromyographic records from 10 adults while reading silently and by Edfeldt (22) who secured similar records from 84 students in the University of Stockholm. On the basis of his broader findings, Edfeldt concluded that "silent speech occurs in the reading of all persons". He found that "good readers engage in less silent speech than do poor readers" and that "the reading of an easy text results in less silent speech than does the reading of a difficult one". The records also supported the view that "a clear text results in less silent speech than does the reading of a blurred one". On the basis of the evidence secured, he concluded that silent speech during silent reading is

a symptom of difficulty in reading. No definite evidence was secured concerning its relation to comprehension "although it appears likely that it may do so". "In any case, it seems quite clear that all kinds of training aimed at removing silent speech should be discarded".

Speed of oral and of silent reading. Previous studies have shown that many factors influence speed of reading, particularly silent reading. Winchester and Gibbons (116) carried on a study among 40 normally hearing veterans to determine the influence of a concurrent noise on "rate" of oral reading. The data secured showed that auditory disturbances "not to exceed 80 db do not produce clinically measurable alterations in oral reading rate". Caution was urged in applying these findings to cases of hearing impairment. Gibbons, Winchester and Krebs (35) secured data from 60 normally hearing veterans to test the hypothesis "that sustained reading, in and of itself, may have an influence upon oral reading rate". A passage of 2000 syllables was read with headphones in an acoustically isolated room. Stop watch records were secured at the end of every 200 syllable unit. No statistically significant differences from unit to unit occurred, indicating that sustained reading does not produce variations in rate of reading within the conditions existing in this study.

Factors influencing speed of silent reading were also considered. Letson (63) compared the scores of 601 college freshmen when reading two passages that differed in difficulty for the same purpose and two passages of equal difficulty for different purposes. He found that "difficulty of the material exerts a greater influence on rate than does purpose" (reading for the story and for complete mastery of ideas and details) and that "slowing down to read more difficult material is important to good comprehension" but is not necessary for mastery of the material. The last statement is somewhat at variance with earlier findings and merits additional study.

Marvel (72) secured data from high school pupils to compare "the effect of the tachistoscope and the effect of 'set' for speed on the improvement of rate and comprehension in reading". He found that "verbal set emerged as the most significant manipulative variable" and seemed to improve speed of reading more uniformly than did the tachistoscope. Comprehension did not improve, however. If the pupil is properly motivated to increase speed of reading, "he will evolve a technique suitable to himself by means of which his reading will improve". Previous research shows also that appropriate training will facilitate progress.

Relationship between speed and comprehension in silent reading. Using the data secured in the study (63) reported earlier, Letson (62) studied the relationship between speed and comprehension in reading. He found that "the relationship between speed and comprehension is high when comprehension is the number of right responses" but "nega-

tive and low when comprehension is the ratio of right responses and the number attempted". The relationship is high for easy material "but decreases as the difficulty of the material increases". He also reported that "a reader tends to maintain a reading rate independent of difficulty of material or of purpose". This finding is somewhat at variance with the results of previous research.

Studies of interpretation in reading. In a summary of previous studies Gray (41) pointed out the fact that methods of approaching the study of interpretation have varied widely over the years. Wide variation is further illustrated in research reported during the past year. In a study among 47 college students Thayer and Pronko (102) presented an ambiguous passage to determine the effect of an ascribed source on its interpretation. The responses to questions following the reading of a passage showed that "the intimation of the source of the passage influenced the readers' structuring of it along the general lines of the conceptual stereotypes attached to the magazine or the type of magazine represented".

Thayer and Pronko (101) analyzed answers of 112 college students to questions following the reading of each of five fictional excerpts and their checking of a personality profile chart containing 25 items relating to the central character in each excerpt. Of great pertinence is the fact that none of the characters were described in the excerpts. The reported clearness of mental images varied from 59 percent of the students, "who saw the third character clearly to 91 percent who reported seeing the second character clearly". Not all the characters were equally liked, the variation being from 80 percent of the students who liked the first character to 20 percent who liked the second character. From all the evidence secured the investigators concluded that to the degree a reader's "psychological history is unique and specific to him" he will "react individually to the reading matter and its references". "To the degree that our subjects had common backgrounds with the language and experiences 'tapped' by the reading excerpts, to that degree did they show a concurrence in responses". A final conclusion was that the experiences which "the reader brings to his reading are more important than the black marks on white paper".

Richardson and Church (83) compared the meanings of seven proverbs as reported by 64 children, ages 7.10 to 12.5, and to those of 30 normal adults, ages 14 to 67. Studies of the responses "revealed significant developmental changes in orientation to the task as one of definition rather than of evaluation, explanation, conundrum-solving, free association, etc.; in interpretation of the proverbs as general statements instead of concrete representations; and in comprehension of the metaphorical nature of the proverbs and ability to coordinate the two levels of meaning". As the increasing complexity of the task is appreciated new obstacles to good achievement arise, so that deficiencies are indicators of increasing maturity.

Critical reading. Because of its great significance in current life, increasing attention has been given to the nature and development of critical reading. As indicated earlier, both Sochor (93) and Eller and Dykstra (23) presented summaries of pertinent research during the year. The latter related specifically to the reader predisposition factor which the investigators warned cannot be ignored.

Maney (71) secured scores on various tests to determine relationships of critical reading comprehension of science material to various other abilities. The data showed "a substantial relationship between literal and critical reading comprehension of science material"; a high relationship between verbal intelligence and proficiency in literal comprehension of science material and a substantial relationship in the case of critical reading, and a high relationship between "general" reading comprehension and literal reading ability of science material and a substantial relationship between "general reading" comprehension and critical reading comprehension of science material. On the basis of the evidence secured, she concluded that critical reading comprehension in science is a complex of skills and abilities which "may be predicted with a fair degree of accuracy from scores on group tests of verbal intelligence and 'general' reading tests".

Sochor (92) reported the results of a study similar to that by Maney (71) but relating to the social studies. She found relationships similar to those cited above for science; also that "reading comprehension in social studies appears to be a composite of many skills and abilities which apparently function at various levels of mental ability".

Reading interests. The fact that children and youth fail to engage in personal reading in harmony with the hopes and expectations of parents and teachers has stimulated continuous efforts over the years to secure through research a better understanding of the problems faced in developing desirable reading interests. In a study involving 1800 children in grades four to six in some schools in Kentucky and Tennessee, Wang (110) found that at least ten percent of the children in the upper and middle groups in intelligence included reading among their preferred recreational activities. Those in the lowest mental group did not.

The Institute of Student Opinion (51) carried on an inquiry early in 1959 among 10,149 teen-age pupils representing all areas of the country concerning their reading interests and habits. The responses indicated that 23 percent of them had read comic books on the previous day; 67 percent had read articles or stories in magazines (not comics), the modal amount of time being from 15 to 30 minutes; 78 percent had read newspapers, devoting about the same amount of time to it as to magazine reading, and 64 percent were currently reading a book not included in school reading. The chief source of books for personal reading was the school library. The modal amount of time spent listening to TV was from two to four hours.

A study by Slover (91) among 365 fourth-grade pupils showed that 90 percent read comic books. The most frequent reasons given were: "funny", "interesting", "like them better", "pictures". These findings and those reported above support Witty's findings to the effect that comic books are most popular about the third- and fourth-grade levels. In a study of the pupils in twelve seventh-grade classes Blakely (10) failed to find evidence of significant differences between comic book and non-comic book readers in intelligence, reading ability, school achievement or behavioral traits. Furthermore, comic book readers "read more library books" than those who do not read comic books.

In a study of the professional reading among 50 elementary school teachers Fisher (29) found that each teacher represented a complex reading pattern "to be understood only through detailed case-study descriptions". The teachers reporting varied all the way from avid readers to those who read very little. Some of the most important factors stimulating professional reading were interest, teaching experience, encouragement of the school principal, and sense of duty.

Factors associated with reading deficiency. The search for personal factors associated with reading deficiency continues unabated. Hermann (48) secured evidence which led him to conclude that "word-blind persons more frequently than normal readers show uncertainty in the discrimination between right and left". He conceives word-blindness not as "an isolated suffering with dyslexia-dysgraphia; but that it is one side of a more general disturbance in the symbol functions. There is underdevelopment of the function: direction in space, transmitted by dominant genes".

Eames (21) compared 24 reading failures having endocrine dysfunction with 100 reading failures without it and with 100 controls. He concluded that "the commonest endocrine disorder likely to be found among reading failures is hypothyroidism", that pupils having endocrine difficulties "will tend to score lower than other reading failures" on reading tests, and "are likely to exhibit more emotional difficulties related to reading". Stott (97) identified pupils who had spent at least two weeks in a hospital during their first two years of life and matched them for comparison against controls. He found that the hospitalized group was significantly more retarded, 27.5 percent being "non-readers or definitely backward compared with 11.3 percent of the controls".

Richardson (82) compared 97 pairs of successful and retarded readers in the third, fourth and fifth grades to determine group differences in the incidence of physical factors. He found far more physical difficulties or deficiencies among the retarded readers. However, the successful readers had also had many physical difficulties. The significant finding was that all the successful readers "were emotionally well adjusted" and "came from stable and secure home backgrounds". "Whether or not a

physical anomaly led to school failure and maladjustment depended upon home conditions".

Steinbaum and Kurk (95) compared the visual performance of a fourth- and a fifth-grade group with greater or less retardation in reading. In harmony with earlier findings, group averages failed to reveal significant differences; however, there was a slight visual advantage for the better reading group. The fact was pointed out that this does not negate the importance of good vision in reading. Eames (19) compared the extent of brain damage among reading failures and non-failures. The data secured showed that brain-damaged children who fail to learn to read "exhibit more muscular imbalance" and "more lateral dominance variations" than similarly afflicted children who learn to read.

Tabarlet (99) compared the average and retarded readers in the fifth grade of 29 white elementary schools in Baton Rouge and found that "poor mental health", as indicated by Mental Health Analysis scores, "and reading retardation go together". No cause-effect relationships could be assumed on the basis of the evidence secured. Graham and Kamano (40) secured data from two groups to test the hypothesis that "the 'criminal' unsuccessful reader would produce a psychopathic profile (as measured by the Wechsler-Bellevue Intelligence Scale) but that the 'criminal' successful reader would not". The former group was found to be inferior both to Wechsler's standardization group and to the latter group "in verbal sub-tests and Digit Symbol".

Readability. The practical importance of valid measures of reading ability has stimulated continued search for the basic factors involved. To this end Stolurow and Newman (96) made a factor analysis by the principal axis method, the factors being rotated by the Quartimax Method, of "an intercorrelation matrix originally prepared by Gray and Leary reduced from 44 to 23 'variables'". "After rotation, two factors, tentatively described as an easy vs. difficult word and a difficult vs. easy sentence factor were found to account for approximately 30 per cent and 20 percent, respectively, of the accounted for variance". These two factors plus eight additional ones accounted "for approximately 93 percent of the total variance", each of the eight accounting for from 8.2 to 2.5 percent of the variance.

Brinton and Danielson (16) also made a factor analysis of twenty language elements used in the Gray-Leary study to identify underlying relationships that might suggest "theoretical bases of readability". According to their calculations, six factors were identified, two of which "appear to be quite clear from a consideration of elements". Factor I was called a vocabulary factor ("word frequency", "word length" and "word familiarity"). The elements in Factor II related mainly to sentences ("average sentence length in words", "average sentence length in syllables" had positive loading; "number of simple sentences", and "number of explicit sentences" had high negative loadings). The

investigators indicated that through further research the nature of some of the other factors they hypothesized can be clarified.

Recognizing the importance of vocabulary in measurements of readability, McCracken (69) carried on an experiment with 124 fifth- and sixth-grade pupils "to determine the effect made upon comprehension scores--by changing difficulty of the selections read". On the basis of the data secured, he concluded that writing passages in harmony with a set of vocabulary standards "may not increase their readability as much as indicated". Questions arise as to whether all other factors of readability were held constant.

Levy (65) attempted to produce a more readable modification of the Allport Vernon Study of Values by making it less verbally complex. The two forms were tried out with low, high and medium vocabulary groups. The high vocabulary group did equally well on the two forms but the low vocabulary group made significantly different scores on the two forms. It was concluded that for this group the revision provided a more valid test of value orientation. Much work "remains to be done" before the modified form can be accepted as a valid equivalent for all subjects.

Bloomer (12) analyzed samples from each of 23 books, Grades I-VI, to test the hypothesis "that reading difficulty is related to the modifier load, the length of modifiers, and the discriminability of the modifiers in a reading passage". He found that each of these elements increased with grade level, and that "modifiers become more difficult to discriminate as the grade increases." He also pointed out that the reader is less free to apply his own referents as the number of modifiers increases.

Those interested in suggestions for further research concerning readability are referred to the article by Powers and Kears (79) which summarizes "the authors' opinions as to some of the needed areas of further research--as mapped out during a University of Wisconsin Ph. D. project." Powers and Ross (80) also present calculating diagrams on recently "recalculated readability formulas originated by Flesch, Dale and Chall, Gunning, and Farr, Jenkins and Patterson".

Hygiene of reading. The long and continued contributions of Dr. Miles A. Tinker to this field merit highest commendation. In a study (104) of the length of work periods in visual research he compared the speed of perception in work periods of 1 1/2, 5 and 10 minutes when reading under 5, 25 and 200 foot-candles of light. The results were the same for the three work periods with 5 vs. 25 foot-candles. In the 25 vs. 200 foot-candles comparison there was no difference for the 1 1/2 minute work period but speed of reading was significantly slower for both the 5 and 10-minute periods. It was suggested that glare had a deleterious effect.

In a second study (105) Tinker compared brightness contrast, illumination intensity and visual efficiency. His findings supported the results of pre-

vious studies to the effect that speed of reading is gradually reduced as the brightness contrast between print and paper is reduced with illumination held at 25 foot-candles. "Increasing illumination intensity above 5 foot-candles had no important effect upon speed of reading when the brightness contrast between print and paper was large (0.756)" but did increase speed when lower contrasts between paper and print were used and illumination was not above 25 foot-candles.

The Teaching of Reading

The studies summarized under this heading are concerned with problems relating to reading programs and instruction.

Studies of achievement and progress in reading. Wilson (114) compared the achievement in reading of pupils in a school system in southern California and a British Columbia city. Whereas both cities scored above the national norm in reading on the California Achievement Test, the British Columbia city scored higher on the average. The pointed question raised was why should the teachers of British Columbia with only two years of training secure better results in teaching reading than the California teachers with four years of training? Fridman (32) compared the reading achievement of a Catholic Parochial School in 1940 and 1956 on the Gates Reading Tests. In general the scores were about a year higher in 1956 excepting in rate. In seeking to promote greater comprehension, speed had been sacrificed to some extent.

Ketchum and Laffitte (55) compared the longitudinal records of fifty children in mental and reading ages beginning with a chronological age of 84 months and continuing to 144 months. The data secured confirmed earlier findings to the effect that children differ widely in mental ability and in reading achievement at each age level, that these differences increase with age, and that each child shows a high degree of individuality in his progress.

Hampleman (46) compared the achievement in reading of two groups of 58 pupils at the end of the sixth grade. They were classified into two groups in respect to time of school entrance, those six years and three months of age and under, and those six years four months and older. Whereas the latter group averaged slightly higher the difference was not statistically significant. Further study of the problem is in progress. The findings indicated that the late entrance pupils had not been penalized in respect to gross achievement at the end of the sixth grade. Kinloch (57) reported the status in reading of pupils entering the second grade who received guided work at home during the preceding summer vacation. A comparison of scores showed that the pupils who received such guidance gained two months on the average while the control group lost two months.

Kasdon (52) studied the early reading background of the 50 highest ranking freshmen in reading in nine

colleges. He found that slightly over half had learned to read before entering school, two-thirds of them having been taught to read at home. Twenty percent believed that they became "fast comprehenders" because they read a great deal. Half of the students attributed their interest in reading to curiosity about the content of books.

Reading Readiness. Bremer (15) was interested in the extent to which the Metropolitan Readiness Test predicts success in reading. He compared the scores of 2,069 pupils on this test with their scores on a reading test at the beginning of the second grade. The correlation was .40. Bremer concluded that readiness tests are "far more useful for planning instruction to overcome deficiencies in readiness". By measuring achievement in reading immediately following a summer vacation, the findings may not provide a true measure of the predictive power of the tests. Banham (4) developed a check list of reading readiness which was shown to have value when used carefully along with the results of intelligence and reading readiness tests "in judging when children are reading to profit by lessons in reading". Robinson, Letton, Mozzi and Rosenbloom (85) studied the value of the Childrens Visual Achievement Forms, as a predictor of reading achievement in Grade I. Statistical treatment of data secured from 87 children led to the conclusion that the CVA "is not a good predictor of first-grade reading achievement".

Studies of basal readers and basal reading instruction. McAnarney (68) made a careful analysis of the five most recently adopted primary basal readers in Kansas to determine if they reflected the results of their adoption periods. He found marked discrepancies between the practices followed and research findings. He urged that those responsible for adoptions should examine the readers under consideration more carefully in the light of the findings of research.

Williams (113) identified 33 critical reading skills that were emphasized in ten basic reading series. When the provisions for critical reading in each series was checked against this list it was found that only three of the skills were stressed in all the series. The largest number found in a given series was 24 and the smallest number was 11. The fact was emphasized that training in critical reading should be introduced from the beginning and provided systematically throughout subsequent grades.

Baranyai (5) selected words and phrases in the readers used to determine the appropriateness of the language included for Hungarian children, ages 8 to 10. Through the use of tests she found that from 40 to 52 percent of the expressions selected for testing were not understood. She rightly points out that children of limited mastery of the English language present problems in teaching reading that differ in degree from those of the normal American child. Clyde (18) analyzed eight third-grade readers to identify the amount of information included concerning occupations. She found considerable

pertinent information but as a rule it was not organized and presented to be useful for a study of occupations. No evidence was presented to show that third-grade readers are more appropriate for such materials than books relating to such fields as the social studies.

The effort to provide better for individual differences has raised pointed questions during recent years concerning the extent to which group instruction and basal readers should be used. The results of two "exploratory" rather than rigidly controlled studies which used a combination of group and individualized teaching were reported. Sharpe (89) reported the progress of a second-grade class which received one or more days of group instruction each week, supplemented by individualized teaching and self selection of the materials read. Bohnhorst and Sellers (13) reported the results of a study in the primary grades in which pupils received alternating group and individualized instruction for eight weeks during each semester followed by individualized instruction the next year. Three of the significant findings of these studies were that the two types of instruction may differ in effectiveness among pupils of different reading ability, that teachers vary in their judgments as to relative effectiveness of the two procedures, and that an effective individualized program requires much time and effort.

Phonics. Bear (7) compared the progress of two groups of pupils during the first grade. The same basal reading program was used with each group excepting one group used a synthetic approach and the other the whole-word approach to phonics, which was integrated with the basal reading materials. During the first half of the year "no differences were apparent in the results produced by the two methods". On tests given at the end of the year, the synthetic phonic group scored higher. Tests are planned at the end of three years to provide further evidence concerning the progress of the two groups. They should be repeated again at the end of the sixth grade. Luser, Stanton and Doyle (66) gave 43 drill lessons on phonics to third- and fourth-grade pupils. As compared with control groups which did not receive the drill lessons, the experimental groups did significantly better on standardized tests of oral reading, paragraph meaning and spelling. Both studies support the contention that systematic training in word attack is essential.

Vocabulary. Research during the last two decades has shown that the understanding vocabulary of first-grade children is much larger than was earlier thought to be true. Shibbes (90) presented evidence which lends further support to this view. The size of the child's understanding vocabulary was greater among children of high as compared with low intelligence and among children of monolingual as compared with bilingual background. Horn (49) compared the "vocabulary of everyday writing" with the words in the 1923, 1927 and 1956 editions of the Webster's New International Dictionary.

From the evidence secured he concluded that "new words are appearing in written communication, but the total number and frequencies of these words are small". However, new meanings are being added to existing words "which result in increasing the usability of our everyday language".

Grouping for instruction in reading. The problem of grouping is one of the most challenging faced today in organizing instruction. Clarke (17) made a statistical study of the effect of grouping on variability of achievement using data from 21 third-grade classes in five schools. He found that grouping "by California Reading Test scores reduces variability in California Mental Maturity I. Q. 's by 10 percent" and that "grouping by I. Q. reduces variability in California Reading Test scores by 9 percent". On the basis of his findings, the investigator favored "a reading test over an intelligence test, if the purpose of grouping is to reduce variability in achievement". As found in previous investigations, grouping on the basis of one factor reduces variability in the other factors only slightly.

Aaron, Goodwin and Kent (1) divided the pupils of three fourth-grade rooms into six classes on the basis of reading ability. Each teacher taught two groups, one of which could be taught with materials of fourth-grade difficulty. At the end of six months the teachers concluded "that the cross-class grouping was more efficient in terms of teacher and pupil time and energy than the familiar within-class grouping". In such studies, the effectiveness of teachers' guidance of pupils in reading activities in other subjects should be considered also. Studies of the effectiveness of homogeneous grouping were reported by Hart (47) at the fourth- and fifth-grade levels and by Bernard (9) at the upper-grade level. The findings in both cases were distinctly favorable to such grouping. The fact was pointed out that homogeneous grouping may not be desirable or possible in some schools.

Developmental reading at the college level. Robinson (84) secured responses from 401 secondary-school reading specialists concerning their work and preparation. The responses indicated that they are extending their influence beyond the confines of the English Department and are beginning to help teachers in other curriculum areas. The titles assigned to them do not always accurately designate their duties. Ninety-eight percent of these specialists felt the need of more intensive training in reading. One of the sources of greatest dissatisfaction related to the present structure of reading programs and lack of acceptance of such programs in schools. Patterson (77) presented twelve generalizations to 31 teachers concerning their participation in supplementing the formal reading program. The responses indicated that many techniques are now being used with major emphasis on vocabulary and background development. However, classroom teachers do not feel competent to help pupils in reading due to lack of training in this field.

Tormey and Patterson (106) provided instruction

in reading during the spring semester to students ranking low in reading. The training resulted in increasing the median percentile ranking of ninth-grade pupils in reading from 20 to 49. Corresponding increases were achieved in the tenth, eleventh and twelfth grades. Both teachers and students agreed that the course was too short to achieve best results. A "booster" course was recommended for those who did not profit sufficiently from the first course. Such a proposal fails to recognize the fact that the need for developmental training in reading is continuous throughout the secondary school period.

In order to determine criteria for selecting supplementary reading science books for intellectually gifted students, Barnes, Beck, Reiner and Washton (6) submitted for checking a list of 13 "values of a book" and 10 "qualities of a book" to 150 members of the National Association for Research in Science Teaching. The two highest ranking items in the first list were "it suggests further problems" and "it stimulates further reading". The highest ranking item in the second list was it is "accurate and authoritative". The findings as a whole are illuminating and highly suggestive.

Reading at the college level and in adult life. The fact that many college students fail because of limited ability to read is too well known to merit re-emphasis. After eight years of detailed work with such students, the reading staff at DePaul University, as reported by Halfter and Douglass (45) have concluded that the chief difficulty is not in the basic skills of recognition and comprehension but rather in the thinking skills involved in most reading activities. They estimate that two-thirds of entering students have difficulty for this reason.

Both Lee (61) and Beasley (8) confirm the findings of earlier studies to the effect that the reading ability of college students can be materially improved through appropriate training and that the improvement made is retained after three to six months. The latter found that retention was greatest among those ranking highest in reading ability. In a study of 477 students in Home Economics, Leahy (60) found that whereas they averaged as high in reading as other students on entering college, they tended to regress during subsequent years in college in comparison with national norms. Questions were raised as to the adequacy of the challenge presented through reading by the courses taken.

Added evidence that the reading ability of executives and of those in industry can be improved and that the improvement can be retained was presented by Murdick (75) and by Sullivan (98). Machines were used in the first study but not in the second.

Remedial reading. Bliesmer (11) reported an exploratory study of three methods of measuring the progress of retarded readers: gain from beginning to end of the training program; gain during training program as compared with average yearly gains during previous years; and "differences between potential-achievement gaps at the beginning and end

of the clinic year". The advantages and limitations of each procedure were considered, the final conclusion favoring the last of the three methods.

In a study of 1,083 third-grade pupils, Wilson (115) found that all who scored at the tenth, or lower, and at the fiftieth percentile on the California Test of Personality ranked about equally low in achievement in spelling, arithmetic, reading and intelligence. Of special significance is his conclusion to the effect that efforts in the early grades to produce higher levels of academic achievement than are normally made by this group appear to be beneficial rather than harmful to their personality structures. He also pointed out that the kinds of pressure exerted among groups and individuals merit careful study.

Evidence of the value of appropriate remedial instruction was reported by Friedmann (33) who carried on a study in England. The rate of progress during the training period was three times as rapid as that made previously. Turner (107) reported that such instruction "pays dividends at the high-school level" not only in improved reading ability but also in loss of fear when engaging in reading activities, particularly oral reading. Braddock and Kraus (14) working with remedial readers found that emphasis on vocabulary growth in the various courses taken was beneficial. Those interested in the nature of the clinical services in New York City, and in the types and causes of reading deficiency identified, are referred to the article by Fite and Mosher (31).

Tests. The need for better tests has resulted both in the critical evaluation of existing tests and in the development of new ones. Levinson (64) analyzed scores made by 776 pupils on the Revised Stanford Binet Scale and found that the vocabulary section favors native-born children. North (76) reported relatively high correlations between scores on the Step Listening Test and various reading tests. Keats (53) compared the difficulty of six vocabulary tests among 600 New Zealand children nine years of age. Woodcock (120) describes the nature, development and evaluation of a test for selecting remedial readers composed entirely of pictures and non-letter symbols.

BIBLIOGRAPHY

1. Aaron, I. E., Goodwin, Frances, and Kent, Vada. "Fourth Grade Teachers Experiment with Cross-Class Grouping for Reading Instruction," *Elementary English*, XXXVI (May, 1959), 305-307.
Reports the progress achieved during six months of instruction in which all the pupils in three fourth-grade rooms were divided into six reading groups on the basis of their reading ability, each teacher teaching two groups.
2. Artley, A. Sterl. "Critical Reading in the Content Areas," *Elementary English*, XXXVI (February, 1959), 122-30.

Reports the results of documentary research in an effort to define the "nature of critical reading", its "relation to literal reading in the content fields", "the factors that predispose the reader to do critical reading, the "basic abilities" involved, and the levels and areas in which it should be developed.

3. Ausubel, David P. and Blake, Elias, Jr. "Proactive Inhibition in the Forgetting of Meaningful School Material," Journal of Educational Research, LII (December, 1958), 145-49.

Reports the results of an experiment at the college level to determine the effect on retention of reading material related to Buddhism as contrasted to the reading of material including all the content of the first passage plus a discussion of similarities and differences between Buddhism and Christianity.

4. Banham, Katharine M. "Maturity Level for Reading Readiness," Educational and Psychological Measurement, XVIII (Summer, 1958), 371-75.

Presents a check list for maturity level for reading readiness to be used as a supplement to reading readiness tests, discusses its validity and the correlations of scores made on it with those on tests of intelligence and reading ability.

5. Baranyai, Elizabeth H. "Verbal Comprehension in Hungarian Children of 8-10 Years," British Journal of Educational Psychology, XXVIII (November, 1958), 262-65.

Reports the findings of a study which aimed, first, to identify children's understanding of words and phrases selected from their school readers and, second, to determine the appropriateness of the readers for the Hungarian children tested.

6. Barnes, Cyrus W., Beck, Alfred D., Reiner, William B., Washton, Nathan S. "Criteria for Selecting Supplementary Reading Science Books for Intellectually Gifted High School Students," Science Education, XLII (April, 1958), 215-18.

Presents the percentages of responses from 150 members of the National Association for Research in Science Teaching ascribing degrees of importance to 13 criteria relating to the "effect on the reader" of a book and to 10 criteria relating to the "qualities of the book".

7. Bear, David E. "Phonics for First Grade: A Comparison of Two Methods," Elementary School Journal, LIX (April, 1959), 394-402.

Reports the results of a controlled experiment in seven classes in the first grade to determine the comparative effectiveness of two methods of teaching reading which differed in the nature and method of phonics instruction.

8. Beasley, Charles E., Jr. "A Freshman Reading Program," Journal of Developmental Reading, II (Winter, 1959), 23-29.

Reports the gain and permanence of gain

after three months of instruction, involving use of mechanical and non-mechanical devices, given to 144 students during thirty-six class hours, supplemented by twenty-four hours of individual and small-group help.

9. Bernard, Mary J. "Homogeneous Grouping for Reading Instruction: Upper Grade Rooms Divided for Reading Lessons," The Chicago Schools Journal, XL (December, 1958), 135-39.

Reports the results of a year of reading instruction given to nine classes which had been classified into homogeneous groups on the basis of the Chicago Reading Test.

10. Blakely, W. Paul. "A Study of Seventh Grade Children's Reading of Comic Books as Related to Certain Other Variables," Journal of Genetic Psychology, XCIII (December, 1958), 291-301.

Summarizes the data secured in a study involving pupils in 12 seventh-grade classes to determine differences in respect to certain variables among pupils differing in the character and amount of comic book reading.

11. Bliesmer, Emery P. "Methods of Evaluating Progress of Retarded Readers in Remedial Reading Programs," The 15th Yearbook of the National Council on Measurements Used in Education, 1958, pp. 128-34.

Compares the results of three methods of evaluating progress of retarded readers in remedial reading programs in the case of two groups of pupils in grades four to eight, inclusive.

12. Bloomer, Richard H. "Level of Abstraction as a Function of Modifier Load," Journal of Educational Research, LII (March, 1959), 269-72.

Analyzes samples from each of 23 books, grades I-VI, to test the hypothesis "that reading difficulty is related to the modifier load, the length of modifiers, and the discriminability of the modifiers in a reading passage."

13. Bohnhorst, Ben A. and Sellers, Sophia N. "Individual Reading Instruction vs. Basal Textbook Instruction: Some Tentative Explorations," Elementary English, XXXVI (March, 1959), 185-90, 202.

Reports an "exploratory" rather than "rigorously experimental" study in two classrooms in the primary grades in which the pupils received alternating group and individualized instruction for eight weeks during each of two semesters, followed by individualized instruction the next year.

14. Braddock, Richard and Kraus, Sidney, "An Experiment in Accelerating the Vocabulary Growth of Remedial Students," College Composition and Communication, IX (December, 1958), 216-19.

Reports the results of efforts in a freshman communication course for low ranking students to increase their meaning vocabulary in the

various courses taken.

15. Bremer, Neville. "Do Readiness Tests Predict Success in Reading?" Elementary School Journal, LIX (January, 1959), 222-24.

Presents Pearson product-moment coefficient of correlation of the scores of 2,069 pupils on the Metropolitan Readiness Tests, Form R, given during the first month of the first grade and on the Gray-Votaw-Rogers General Achievement Tests, Primary Test, Form Q, given at the beginning of the second grade.

16. Brinton, James E. and Danielson, Wayne A. "A Factor Analysis of Language Elements Affecting Readability," Journalism Quarterly, XXXV (Fall, 1958), 420-26.

Presents the results of a factor analysis of twenty language elements to identify underlying relationships that might suggest "theoretical bases of readability".

17. Clarke, S. C. T. "The Effect of Grouping on Variability of Achievement at the Grade III Level," The Alberta Journal of Educational Research, IV (September, 1958), 162-71.

A statistical study of the effect on variability in achievement of grouping pupils in 21 Grade III classes in five schools on the basis of scores made on the California Mental Maturity Test and on the California Reading Test.

18. Clyse, Juanita. "What Do Basic Readers Teach About Jobs," Elementary School Journal, LIX (May, 1959), 456-60.

Presents the results of an analysis of the selections on eight basic readers for use at the third-grade level to determine the nature and extent of the information given on occupations "as a means of earning a living".

19. Eames, Thomas H., M.D. "Reading Failures and Non-Failures in Children with Brain Damage," American Journal of Ophthalmology, XLVII (January, 1957), 74-77.

Compares the per cent of cases showing evidence of brain damage among reading failures and non-failures.

20. Eames, Thomas H., M.D. "Visual Handicaps to Reading," Journal of Education, CXLI, (February, 1959), 3-35.

Summarizes the comparative eye condition of reading failures and unselected groups of school children examined by the author over a long period of years.

21. Eames, Thomas H., M.D. "The Effect of Endocrine Disorders on Reading," The Reading Teacher, XII (April, 1959), 263-65.

Reports the results of comparisons of 24 reading failures with endocrine dysfunction with 100 reading failures without endocrine dysfunction and 100 controls.

22. Edfelt, Ake W. Silent Speech and Silent Reading. Stockholm, Sweden: Almquist and Wiksell, 1959.

Analyzes electromyographically secured silent speech records while reading of 84 freshman students in the University of Stockholm to

determine the relationship between the occurrence of silent speech in reading and reading ability.

23. Eller, William and Dykstra, Robert. "Persuasion and Personality: Readers' Predispositions as a Factor in Critical Reading," Elementary English, XXXVI (March, 1959), 191-97, 202.

Presents a critical review of the published research relating to critical reading, with special reference to the "reader-predisposition factor," including personal, social and cultural aspects.

24. Ellison, Jerome and Gosser, Franklin T. "Non-Fiction Magazine Articles: A Content Analysis Study," Journalism Quarterly, XXXVI (Winter, 1959), 27-34.

Presents the results of content analyses of nine non-fiction magazines for three-month periods in 1947 and 1957 to identify changes in subject matter, writing style and article length for individual magazines and the group.

25. Engel, Gerald, O'Shea, Harriet E. and Mendenhall, John H., "Projective Responses to a News Article: A Study in Aspects of Bias," Journal of Psychology, XLVI (October, 1958), 309-17.

Reports the results of an experiment, including 233 college students and adults belonging to 8 different groups, to test the hypotheses that "people read and interpret statements dissimilarly" and that "the direction of interpretation of ambiguous reading material is determined by the existing bias of the individual."

26. Faaborg-Andersen, K. and Edfeldt, A. W. "Electromyography of Intrinsic and Extrinsic Laryngeal Muscles During Silent Speech: Correlation with Reading Activity," Acta otolaryng., XLIX (November-December, 1958), 478-82.

Describes the methods used with ten adult volunteers, presents reproductions of the records secured, and compares the extent of electrical activity in several speaking and reading tasks.

27. Fay, Leo C. Improving the Teaching of Reading by Teacher Experimentation. Bulletin of the Indiana University School of Education, Vol. 34, No. 5, September, 1958.

Describes the nature, procedures and results of eleven teacher-designed experiments relating to "word study-phonics", "reading in the content fields", "extensive reading", "helping the handicapped reader", "reading in the high school."

28. Fennema, Elizabeth H. "Mental Imagery and the Reading Process," Elementary School Journal, LIX (February, 1959), 286-89.

Reports the results of a study with beginning fourth-grade pupils to determine the nature and extent of the mental imagery accompanying their reading and the correlation of mental-imagery scores with intelligence and reading a-

- bility.
29. Fisher, Helen. "Teacher Differences in Professional Reading," Educational Administration and Supervision, XLIV (September, 1958), 282-89.
Summarizes the results of information secured from 50 elementary-school teachers to determine their patterns of professional reading and the factors that influenced their reading.
 30. Fishman, Joshua A. and Fishman, Cele Schweid. "Separatism and Integrationism: A Social-Psychological Analysis of Editorial Content in New York Newspapers of Three American Minority Groups," Genetic Psychology Monographs, LIX (May, 1959), 219-61.
Compares the editorial content of five foreign-language newspapers published in New York City to determine the proportion of editorials devoted to their own ethnic group problems and to "separatism-retentionism" on the part of three American minority groups.
 31. Fite, Margaretta W. and Mosher, Margaret M. "The Special Reading Services of the New York Board of Education," The Reading Teacher, XII (February, 1959), 181-86.
Describes the clinical programs and summarizes evidence concerning "symptoms of disturbance", "problems within the family", "situational factors and early school experiences" and "adverse physical developmental findings".
 32. Fridian, Sister Mary. "Reading Achievement in a Catholic Parochial School," School and Society, LXXXVI (November 8, 1958), 403-5.
Compares the achievement in reading of pupils in Grades I to VII inclusive in 1940 and in 1956, as measured by the Gates reading tests for the appropriate grades.
 33. Friedman, S. "A Report on Progress in an L. E. A. Remedial Reading Class," British Journal of Educational Psychology, XXVIII (November, 1958), 258-261.
Compares "the progress in reading of retarded children before entering a remedial class, and after a period of instruction in the class."
 34. Garver, Richard A. "The Labor Press as a Leadership Tool," Journalism Quarterly, XXXV (Summer, 1958), 324-32.
Presents a content analysis of U. S. Labor publications issued during 1956 from headquarters of 34 international trade unions chosen randomly from a possible total of 163 such unions, "to determine the basic purpose of such publications."
 35. Gibbons, Edward W., Winchester, Richard A. and Krebs, Donald F. "Variability of Oral Reading Rate," Journal of Speech and Hearing Disorders, XXIII (November, 1958), 591-93.
Reports and interprets evidence from 60 veterans which was secured to test "the hypothesis that sustained reading, in and of itself, may have an influence upon oral reading rate."
 36. Gilbert, Luther C. "Influence of Interfering Stimuli on Perception of Meaningful Material," California Journal of Educational Research, X (January, 1959), 15-23.
Summarizes data secured from 68 upper division and graduate students "to determine the amount of interference which resulted when a competing stimulus was thrown on the same group of rods and cones immediately following or preceding the presentation of words or phrases of sense material".
 37. Gilbert, Luther C. "Saccadic Movements as a Factor in Visual Perception in Reading," Journal of Educational Psychology, L (January, 1959), 15-19.
Summarizes data from 76 college and graduate students which were secured in exploring "the relationship between the speed and accuracy of perception" in reading simple prose "both with and without saccadic movements of the eyes."
 38. Gilbert, Luther C. "Speed of Processing Visual Stimuli and its Relation to Reading," Journal of Educational Psychology, L (January, 1959), 8-14.
Summarizes data from 64 college and graduate students which were secured to find out the "influence of varying the processing times for the first stimulus before the Ss were permitted to encounter an interfering stimulus."
 39. Gilbert, Luther C. "Genetic Study of Eye-Movements in Reading," Elementary School Journal, LIX (March, 1959), 328-35.
Compares the number of fixations and regressions per 100 words and the duration of fixations in prose reading of 22 subjects when in the second grade and again when college juniors and seniors; also for 7 subjects at the second, fourth and sixth grades and in college.
 40. Graham, E. E. and Kamano, D. "Reading Failure as a Factor in the WAIS Subtest Patterns of Youthful Offenders," Journal of Clinical Psychology, XIV (July, 1958), 302-5.
Presents evidence secured to test the hypothesis that "the 'criminal' unsuccessful reader would produce 'psychopathic' profile but that the 'criminal' successful reader would not".
 41. Gray, William S., "New Approaches to the Study of Interpretation in Reading," Journal of Educational Research, LII (October, 1958), 65-67.
Traces briefly the history of research concerning the interpretation of what is read with special emphasis on the introspective techniques used during recent years.
 42. Gray, William S. "Summary of Reading Investigations, July 1, 1957 to June 30, 1958," Journal of Educational Research, LII (February, 1959), 203-21.
Summarizes major findings of 118 investigations relating to reading; includes annotated

bibliographical references.

43. Haberland, John A. "A Comparison of Listening Tests with Standardized Tests," Journal of Educational Research, LII (April, 1959), 299-302.

Compares the scores made on three listening tests by 110 freshmen college students who were classified into two groups on the basis of reading ability--one above and the other below average.

44. Hage, Dean S. and Stroud, James B. "Reading Proficiency and Intelligence Scores, Verbal and Nonverbal," Journal of Educational Research, LII (March, 1959), 258-62.

Reports correlations in the case of 800 ninth-grade pupils in ten public schools, using scores on the Lorge-Thorndike Intelligence Tests, the Pressey Rate and Comprehension Tests and the Iowa Tests of Basic Skills.

45. Halfter, Irma T. and Douglas, Frances M. "'Inadequate' College Readers," Journal of Developmental Reading, I (Summer, 1958), 42-53.

Compares the scores made by high-school graduates on a battery of entrance tests, including the Survey Section, Diagnostic Reading Test, Form D., with their subsequent grade-point averages to identify the predictive value of the tests used.

46. Hampleman, Richard S. "A Study of the Comparative Reading Achievements of Early and Late School Starters," Elementary English, XXXVI (May, 1959), 331-34.

Compares the mean achievement scores of 58 pupils at the end of the sixth grade, classified into two groups--those six years three months of age and under, and those six years and four months of age and older.

47. Hart, Richard H. "The Effectiveness of an Approach to the Problem of Varying Abilities in Teacher Reading," Journal of Educational Research, LII (February, 1959), 228-31.

Reports the results of a study of the fourth and fifth-grade levels to test the hypothesis that achievement in reading is significantly greater when pupils are grouped homogeneously according to ability than when grouped heterogeneously, "all other factors being equal".

48. Herman, Knud. "Congenital Word-Blindness: (Poor Readers in the Light of Gerstmann's Syndrome)," Acta Psychiatrica and Neurologica Scandinavica, Supplement 108. Copenhagen, Denmark; Munksgaard, 1956.

Presents evidence that the fundamental disturbance in the case of congenital word-blindness is the same as that in the case of Gerstmann's Syndrome.

49. Horn, Thomas D. "What's New in Words These Days?," Education, LXXIX (December, 1958), 203-5.

Compares the vocabulary of "modern everyday writing", as revealed through twelve stud-

ies, with the words in the 1923, 1927 and the 1956 editions of Webster's New International Dictionary.

50. Huelsman, Charles B. "Some Recent Research on Visual Problems in Reading," American Journal of Optometry and Archives of American Academy of Optometry, XXXV (November, 1958), 559-64.

Summarizes pertinent research since 1960 under three headings; "outline form perception", "use of the tachistoscope", and "general visual achievement".

51. Institute of Student Opinion. Teen-Age Reading Habits. Scholastic Magazines, Inc. New York, 33 West 42nd Street, 1959.

Summarizes the responses of 10,149 teenage students from all parts of the country concerning 13 issues relating to their reading interests and habits, secured during February-March, 1959.

52. Kasdon, Lawrence M. "Early Reading Background of Some Superior Readers Among College Freshmen," Journal of Educational Research, LII (December, 1958), 151-53.

Summarizes the results of inquiries among 50 of the highest ranking freshmen in nine colleges in the Los Angeles area on the Speed of Comprehension Scale of the Cooperative English Test, Form Y, to determine if they learned to read before school entrance, how they learned to read, and how they accounted for the fact that they became superior readers.

53. Keats, J. A. et. al. "The Difficulty of Available Word Recognition Tests for Queensland Children," The Australian Journal of Education, II (November, 1958), 168-70.

Reports the results of an evaluation of six vocabulary tests with respect to difficulty, using as a "criterion of difficulty" the number of pupils in a group of 600 approximately nine years old who in a previous study had used a particular word appearing in any of the six tests.

54. Kerrick, Jean S. "News Pictures, Captions, and the Point of Resolution," Journalism Quarterly, XXXVI (Spring, 1959), 183-88.

Summarizes the responses of two groups of 50 college women "to explore the effect (on meaning) of combining pictures and captions", using five pictures and a pair of captions, "one intended to be positive" and "one intended to be negative".

55. Ketchum, Warren A. and Laffitte, Rondeau G., Jr. "How Well are They Learning?" Educational Leadership, XVI (March, 1959), 337-41.

Compares the longitudinal records of fifty children in mental and reading ages beginning with a chronological age of 84 months and continuing to 144 months.

56. King, W. H. "An Experimental Investigation into the Relative Merits of Listening and Reading Comprehension for Boys and Girls of Pri-

mary School Age," British Journal of Educational Psychology, XXIX (February, 1959), 42-49.

Compares the scores of 225 boys and 250 girls on tests of visual and auditory comprehension and relates these scores with the results of a verbal intelligence test and tests of (a) ability to follow directions, (b) immediate recall, (c) word meanings and (d) attention.

57. Kinloch, Florence "Do Children Forget During the Summer?" School and Community, XLV (May, 1959), 9; 40.

Presents the results of a controlled experiment with two groups of 14 pupils each to determine (1) if pupils decrease in reading ability during the summer following first grade training and (2) the influence of guided work during the summer on their status in reading at the end of the summer.

58. Kittell, Jack E. "Bilingualism and Language--Non-Language Intelligence Scores of Third-Grade Children," Journal of Educational Research, LII (March, 1959), 263-68.

Compares the scores of 83 third-grade pupils attending the same elementary school, 42 coming from bilingual environments and the remainder from a unilingual (English) environment, on the California Short Form Mental Maturity Test and the California Reading Test, Primary, Form AA, to determine the influence of bilingualism on intelligence scores and reading achievement.

59. Larson, Robert E. and Selland, Cynthia T., "A Comparison on Reading Ages with Mental Ages," Journal of Educational Research, LII (October, 1958), 55-59.

Compares the scores made on the Kuhlmann-Anderson Intelligence Test and the reading section of the Iowa Basic Skills Tests, Part I, by 475 sixth-grade pupils and summarizes reasons for underachievement as reported by teachers.

60. Leahy, Dorothy M. "Reading Ability of College Home Economics Students," California Journal of Educational Research, X (January, 1959), 42-47.

Reports the findings of a study made among 477 women students in Home Economics to find out how well they read, the extent of their vocabulary, how they compared in reading ability with college students in general, and the relation of reading ability to their grade-point averages.

61. Lee, Maurice A. "Results of a College All-Freshmen Reading Improvement Program," Journal of Developmental Reading, II (Autumn, 1958), 20-32.

Reports the aims, organization, administration, the gains made, the retention after six months, and the critical evaluation by participating teachers and students of a two-hour

weekly, one semester, all-freshmen reading improvement program.

62. Letson, Charles T. "Speed and Comprehension in Reading," Journal of Educational Research, LII (October, 1958), 49-53.

Presents finding of a study of relationships between speed and comprehension in which 601 college freshmen were given four tests: two were read for the same purpose but consisting of simple and difficult material; and two consisting of passages of equal difficulty but read for different purposes.

63. Letson, Charles T. "The Relative Influence of Material and Purpose on Reading Rates," Journal of Educational Research, LII (February, 1959), 238-40.

Summarizes the scores made by 601 college freshmen when reading two passages that differed in difficulty for two purposes: (a) as rapidly as possible for the story and (b) for complete mastery of ideas and details.

64. Levinson, Boris M. "Reevaluation of the Revised Stanford-Binet Scale, Form L Vocabulary as a Test of Intelligence for the Kindergarten and Primary School Child," Journal of Genetic Psychology, XCIII (December, 1958), 237-48.

Analyzes the scores made by 776 pupils on the Revised Stanford-Binet Scale, Form L, to determine (a) its value in establishing an index of intelligence as compared with the full-scale Binet and (b) the relative standing of a different sub-culture (foreign birth) on vocabulary and full-scale Binet.

65. Levy, Jerome. "Readability Level and Differential Test Performance, A Language Revision of the Study of Values," Journal of Educational Psychology, XLIX (February, 1958), 6-12.

Presents the results of an attempt to produce a more readable modification of the Allport-Vernon Study of Values that was equivalent to it in all other respects.

66. Luser, Carolyn, Stanton, Eileen, and Doyle, Charles I. "Effect of an Audio-visual Phonics Aid in the Intermediate Grades," Journal of Educational Psychology, XLIX (February, 1958), 28-30.

Reports the results of a controlled experiment involving 214 third- and fourth-grade pupils to determine the effect of 43 drill lessons in phonics on achievement in oral reading, paragraph meaning, and spelling.

67. Lynn, R. "Disparity of Attainment in Reading and Arithmetic," British Journal of Educational Psychology, XXVIII (November, 1958), 277-80.

Replies to criticisms by Reed and Schonfield of a previous article under the same title as above, reviewing the results of relevant research to support his views.

68. McAnarney, Harry. "A Comparison of the Five Most Recent Kansas Basal Primary Reading

Adoptions with the Educational Thought of Each Adoption Period in the Areas of Content, Vocabulary, Word Analysis, Readiness, and Readability," University of Kansas Bulletin of Education, XIII (Spring Issue, May, 1959), 95-100.

Reports the findings of a study directed by three aims: "(1) to examine the research findings of each adoption period; (2) to establish on the basis of the research findings, criteria to use in the evaluation of the adopted reading series; (3) to compare the Kansas adoptions with the research of each of the adoption periods."

69. McCracken, Robert A. "An Experiment with Contrived Readability in Fifth and Sixth Grades," Journal of Educational Research, LII (March, 1959), 277-78.

Reports the results of an experiment involving 124 fifth and sixth-grade pupils "to determine the effect made upon reading comprehension scores... by changing vocabulary difficulty of the selections read".

70. McEvoy, Poynter, "Media-Habit Survey of Indiana Homes," Journalism Quarterly, XXXVI (Winter, 1959), 63-64.

Summarizes the 260 responses mainly from middle class urban families to questionnaires to determine the extent of use of different mass media and preferences for them among various members of families.

71. Maney, Ethel S. "Literal and Critical Reading in Science," Journal of Experimental Education, XXVII (September, 1958), 57-64.

Reports correlations of scores of 513 fifth-grade pupils on the Pinter General Ability Test, Verbal Series, Intermediate, Form A; The Gates Reading Survey, Form 1, Level of Comprehension; and the experimental edition of The Intermediate Reading Test: Science.

72. Marvel, John A. "Acquisition and Retention of Reading Performance on Two Response Dimensions as Related to Set and Tachistoscopic Training," Journal of Educational Research, LII (February, 1959), 232-37.

Reports the results of studies among 156 sophomore high-school pupils divided into four groups of 39 each to compare "the effect of the tachistoscope and the effect of 'set' for speed on the improvement of rate and comprehension in reading".

73. Mehling, Reuben, "Attitude Changing Effect of News and Photo Combinations," Journalism Quarterly, XXXVI (Spring, 1959), 189-98.

Analyzed responses from seven groups of university students ranging in size from 13 to 30 to determine the relative effectiveness in changing personal attitudes of a photo-caption combination and a story alone.

74. Morrison, Ida E. and Perry, Ida F. "Spelling and Reading Relationships with Incidence of Retardation and Acceleration," Journal of Educational Research, LII (February, 1959), 222-27.

Bases conclusions on the results of the California Achievement Tests and the California Short-form Tests of Mental Maturity given over a period of four years to 1,007 pupils in Grades III through VIII in 36 classrooms in five school districts.

75. Murdick, R. G. "Problems in Developing a Profitable Reading Improvement Course for Executives," Journal of Developmental Reading, II, (Spring, 1959), 22-30.

Presents data comparing the reading efficiency of executives at the beginning, and at the end in the case of 15 of them, from 14 to 18 months after the close of a reading improvement course.

76. North, Robert D. "An Evaluation of the Step Listening Test for the Independent School Testing Program," 1958 Achievement Testing Program in Independent Schools and Supplementary Studies, Educational Records Bulletin, No. 72. New York: Educational Records Bureau, July, 1958, 61-67.

Reports correlations of scores in the Step Listening Test, Form G, with time of the reading sections of the Cooperative English Test, Form Y, and the Stanford Achievement Test, Form J, given to pupils in grades 5, 8 and 11.

77. Patterson, Walter G. "The Role of the Classroom Teacher in Helping Students Improve in Reading," Journal of Developmental Reading, I (Summer, 1958), 31-41.

Summarizes the reactions of 31 high-school teachers to twelve generalizations concerning their participation in supplementing the formal reading program; also the practical suggestions made by the same teachers for helping students improve in reading.

78. Peterson, Wilbur, "Is Daily Circulation Keeping Pace with the Nation's Growth?" Journalism Quarterly, XXXVI (Winter, 1959), 12-22.

Compares (a) circulation of daily newspapers with the total United States population, and with that over 20 years of age, for the years 1929 to 1957 and (b) relationship of circulation to number of households and families.

79. Powers, Richard D. and Kearl, Bryant E. "Further Directions for Readability Research," Journalism Quarterly, XXXV (Fall, 1958), 427-32.

Presents "the authors' opinions as to some of the needed areas of further research in readability, as mapped out during a University of Wisconsin Ph.D. project."

80. Powers, Richard D. and Ross, J.E. "New Diagrams for Calculating Readability Scores Rapidly," Journalism Quarterly, XXXVI (Spring, 1959), 177-82.

Presents calculating diagrams based on recently "recalculated readability formulas originated by Flesch, Dale and Chall, Gunning, and Farr, Jenkins and Paterson".

81. Reed, Graham and Schonfield, David. "Disparity of Attainment in Reading And Arithmetic," British Journal of Educational Psychology, XXVIII (November, 1958), 271-76.

Reviews a recent Article by R. Lynn in which the latter maintained that "anxiety is a cause of good reading but interferes with progress in arithmetic", criticizes Lynn's evidence as inconclusive and makes suggestions for further research. A reply by Lynn follows Graham and Schonfield's discussion.

82. Richardson, J. A. "Physical Factors in Reading failure," Australian Journal of Education, II (April, 1958), 1-10.

Presents a general review of related research and compares 97 matched pairs of retarded and successful readers in the third, fourth and fifth grades to determine group differences in the incidence of physical factors and their possible relation to reading failure.

83. Richardson, Claudia and Church, Joseph. "A Developmental Analysis of Proverb Interpretations," Journal of Genetic Psychology, XCIV (June, 1959), 169-179.

Compares and analyzes the meanings of seven proverbs as reported by 64 children, "aged 7:10 to 12:5" and those by 30 normal adults between 14 and 67 years of age.

84. Robison, H. Alan. "The Secondary School Reading Specialist," The Reading Teacher, XII (December, 1958), 103-106.

Summarizes questionnaire responses of 401 secondary school reading specialists in the five largest cities of 41 states and the District of Columbia, concerning their preparation, earnings, and nature of their work.

85. Robinson, Helen M., Letton, Mildred C., Mozzi, Lucille and Rosenbloom, Alfred A. "An Evaluation of the Children's Visual Achievement Forms at Grade I," American Journal of Optometry and Archives of American Academy of Optometry, XXXV (October, 1958); also monograph No. 237.

Reports correlations between scores made by 87 first-grade children on a battery of tests to evaluate the Children's Visual Achievement Forms as a predictor of reading achievement, skill in handwriting and the need for visual examination.

86. Russell, David H. "Some Research on the Impact of Reading," The English Journal, XLVII (October, 1958), 398-413.

Summarizes in popular style the information gleaned from research concerning the impact of reading, organized under four headings: "characteristics of the reading material", "content of ideas", "characteristics of the reader", "overt responses as effects of reading".

87. Sayles, Daniel C. "The Effects of Medial Testing," Journal of Developmental Reading, II (Autumn, 1958), 43-47.

Reports the results of an experiment including 126 college students, both male and female, to determine the effect of "medial testing on final test performance" in a reading improvement course, meeting two fifty-minute periods a week for six weeks.

88. Schonfield, David. "Changes in Immediate Memory as a Function of Age and Meaning," Alberta Journal of Educational Research, V (June, 1959), 112-18.

Compares the memory of lists of ten words as a result of one presentation of the series to determine the effect on retention of age (grades II, IV, IX and first-year university students) and of meaning relationships between the first and sixth words (autonyms, synonyms and homonyms).

89. Sharpe, Maida Wood. "An Individualized Reading Project," Elementary English, XXXV (December, 1958), 507-12.

Reports the progress in reading of a second grade class which received one or more days of group instruction each week, supplemented by individualized teaching and self-selection of the materials read.

90. Shibbes, Burleigh H. "How Many Words Does a First-Grade Child Know?" Elementary English, XXXVI (January, 1959), 42-47.

Reviews the finding of previous studies and reports the results of The Seashore-Eckers on English Recognition Vocabulary Test given to 183 bilingual and mono-lingual pupils in seven elementary schools in Maine.

91. Slover, Vera. "Comic Books vs. Story Books," Elementary English, XXXVI (May, 1959), 319-22.

Summarizes responses of 365 fourth-grade pupils to the questions: "Do you read comic books?" "Which of these do you enjoy most?" --comic books, story books? "Give a reason for your choice in the question above."

92. Socher, E. Elona. "Literal and Critical Reading in Social Studies," Journal of Experimental Education, XXVII (September, 1958), 49-56.

Reports correlations of scores of fifth-grade pupils on the Pinter General Ability Test, Form B, The Gates Reading Survey, Form 1, Level of Comprehension, and the experimental edition of The Intermediate Reading Test: Social Studies.

93. Socher, E. Elona. "The Nature of Critical Reading," Elementary English, XXXVI (January, 1959), 47-58.

Reviews expert opinion and research findings, as presented in 97 articles, reports and texts, under the following headings: "Definitions of Critical Reading", "Complexity of Reading Comprehension", "Thinking: A Basic Process", and "Literal and Critical Reading".

94. Spache, George D. Good Reading for Poor Readers. Champaign, Ill.: The Garrard Press, 1958.

Presents an extended list of books for poor readers, classified under such headings as adventure, with indexes of reading levels and interest levels. Prefaces this list with summaries of research relating to choosing books for children, the right type of book, bibliography and methods of estimating readability.

95. Steinbaum, Milton and Kurk, Mitchell. "Comparison of Visual Performance in Two Classes of Below Average Readers," Journal of the American Optometric Association, XXXV (October, 1956), 194-96.

Compares the visual status of two groups of below average readers, namely, 17 fourth-grade pupils with greater average retardation in reading and 18 fifth-grade pupils with lesser average retardation.

96. Stolurow, Lawrence M. and Newman, J. Robert. "A Fractional Analysis of Objective Features of Printed Language Presumably Related to Reading Difficulty," Journal of Educational Research, LII (March, 1959), 243-51.

Presents the results of a factor analysis by the principle axis method, the factors being rotated by the quartimax method, using as a base an intercorrelation matrix originally prepared by Gray and Leary, the number of variables being reduced from 44 to 23.

97. Stott, D. H. "Infantile Illness and Subsequent Mental and Emotional Development," Journal of Genetic Psychology, XCIV (June, 1959), 233-51.

Examines the school progress and emotional development of all Bristol-born children "who had spent at least two weeks in a hospital between 7 and 11 years ago."

98. Sullivan, Eugene A., Jr. "Some Experiences with Reading Improvement in General Motors," Journal of Developmental Reading, II (Winter, 1959), 12-16.

Presents records of individuals of different personnel divisions, showing percent of gain in reading efficiency and extent of retention after six months, as a result of twenty hours of training without the use of mechanical devices.

99. Tabarlet, B. E. "Poor Readers and Mental Health," Elementary English, XXXV (December, 1958), 522-25.

Compares the scores of retarded and of average readers in the fifth grade of 29 white elementary schools in Baton Rouge on ten items of the Mental Health Analysis, the retarded readers including pupils with an I. Q. of 86 or above and reading scores of 3.5 or below.

100. Taylor, Earl A. "The Fundamental Reading Skill," Journal of Developmental Reading, I (Summer, 1958), 21-30.

Presents and interprets data secured

through eye-movement records showing average achievement of readers by grades in terms of "fixations per 100 words", "regressions per 100 words", "average span of recognition", "average duration of fixation", also "average rate of comprehension".

101. Thayer, Lee O. and Pronko, N. H. "Some Psychological Factors in the Reading of Fiction," Journal of Genetic Psychology, XCIII (September, 1958), 113-17.

Analyzes and interprets the responses of 112 college students to questions and to their checking of a personality profile chart, following the reading of five fictional excerpts, to determine "the relative significance of a wide variety of contributing factors" to the reading of fiction.

102. Thayer, Lee O. and Pronko, H. H. "The Effect of Conceptual Stereotypes in an Ambiguous Situation," Journal of General Psychology, LX (April, 1959), 269-73.

Presents an analysis of responses of 47 college students to an ambiguous fiction passage to determine "the effect of ascribed source" on its interpretation.

103. Tinker, Miles A. "Recent Studies of Eye-Movements in Reading," Psychological Bulletin, LIV (July, 1958), 215-31.

Summarizes significant findings of 72 studies of eye-movements in reading, published from January, 1955 to October, 1957 under the following headings: "techniques of measurement"; "analysis of the reading process"; "training to improve eye movements"; "topography and eye movements"; "eye-movements and fatigue."

104. Tinker, Miles A. "Length of Work Periods in Visual Research," Journal of Applied Psychology, XLII (October, 1958), 343-45.

Compares the relative efficiency in the case of 180 university sophomores during work periods of $1\frac{1}{2}$, 5 and 10 minutes in measuring speed of perception in reading under 5, 25 and 200 foot-candles of light.

105. Tinker, Miles A. "Brightness Contrast, Illumination Intensity and Visual Efficiency," American Journal of Optometry and Archives of American Academy of Science, XXXVI (May, 1959), 221-36.

Summarizes speed of reading by 367 university sophomores of sentences printed with varying brightness contrasts and illumination intensities to determine effect of illuminating intensity upon visual efficiency with variations in brightness contrasts between print and paper.

106. Tormey, Mary K. and Patterson, Walter G. "Developmental Reading and Student Evaluation," Journal of Developmental Reading, II (Winter, 1959), 30-43.

Presents data showing gains made by ninth-

- grade students given reading instruction during a spring semester; also their evaluation of the effectiveness of the training given and suggestions for its improvement.
107. Turner, Carla S. "Remedial Reading Pays Dividends in the Junior High School," English Journal, XLVIII (March, 1959), 136-40, 153.
Reports the progress in vocabulary, comprehension and speed of 40 junior high school pupils with I. Q.'s above 90, this progress resulting from a semester of remedial reading provided individually in groups of from 15 to 20; indicates changes proposed in future training programs.
 108. Vernon, M. D. Backwardness in Reading: A Study of its Nature and Origin. Cambridge: the University Press, 1958.
Summarizes experimental and clinical studies of those who for some reason or other are unable to master the simple mechanics of reading, using the following headings: visual perception, auditory perception, innate factors, acquired defects, environmental factors.
 109. Waldman, John and Triggs, Francis Oralind. "The Measurement of Word Attack Skills," Elementary English, XXXV (November, 1958), 459-63.
Reports the correlations between the scores made by pupils in grades 4, 5, 6 and 7 in word recognition, comprehension, vocabulary and rates of reading, as measured by the Diagnostic Reading Tests: Survey Section, Lower Level, and in Spelling, as measured by the Lincoln Diagnostic Spelling Test, Intermediate Forms A, B, C and D.
 110. Wang, James D. "The Relationship Between Children's Play Interests and their Mental Ability," Journal of Genetic Psychology, XCIII (September, 1958), 119-31.
Reports the results of a study among 1800 children in grades four, five and six in certain schools of Tennessee and Kentucky to determine the preferences of children of varying levels of mental ability for certain play activities, including reading.
 111. Wertheimer, Michael and Gillis, Willie Mae. "Satiation and the Rate of Lapse of Verbal Meaning," Journal of General Psychology, LIX (July, 1958), 79-85.
Analyzes the responses of 501 subjects to three lists of twelve words each, involving six variables, to determine the influence of each on the rate of lapse of meaning.
 112. Westover, Frederick L. "A Comparison of Listening and Reading as a Means of Testing," Journal of Educational Research, LII (September, 1958), 23-26.
Summarizes the results of tests given to 198 college students to determine "the effects of the factors of listening to questions and of reading questions on the students' performance on objective tests".
 113. Williams, Gertrude. "Provisions for Critical Reading in Basic Readers," Elementary English, XXXVI (May, 1959), 323-30.
Presents the results of an analysis of the provision for promoting growth in critical reading in the various readers of ten series for readers from the preprimers to the sixth grade, inclusive.
 114. Wilson, John A. R. "Differences in Achievement Attributable to Different Educational Environments," Journal of Educational Research, LII (November, 1958), 83-93.
Compares time allotted to various school activities in grades I and II, preparation required for teaching, class size, rates of acceleration and retardation, and third-grade achievement scores in reading of pupils in a school system in Southern California and in Alberta, Canada.
 115. Wilson, J. A. R. "Achievement, Intelligence, Age and Promotion Characteristics of Students Scoring at or Below the Tenth Percentile on the California Test of Personality," Journal of Educational Research, LII (April, 1959), 283-92.
Analyzes scores of 1083 third-grade pupils in two school systems to determine "the achievement, age, intelligence and promotion characteristics" of those who scored at or below the tenth percentile on the California Test of Personality.
 116. Winchester, Richard A. and Gibbons, Edward W. "The Effect of Auditory Masking upon Oral Reading Rate," Journal of Speech and Hearing Disorders, XXIII (August, 1958), 250-52.
Presents data from 40 normally hearing veterans to determine "the influence upon oral reading rate provided by a commonly used masking (distracting) noise."
 117. Witty, Paul A. and Sizemore, Robert A. "Studies in Listening, I: Relative Values of Oral and Visual Presentations," Elementary English, XXXV (December, 1958), 583-52.
Summarizes the results of twenty-five studies concerning the relative merits of auditory and visual presentations (listening and reading).
 118. Witty, Paul A. and Sizemore, Robert A. "Studies in Listening, II: Relative Values of Oral and Visual Presentations," Elementary English, XXXVI (January, 1959), 59-70.
Summarizes the results of "studies of the relative value of listening to materials presented in lecture form as compared with reading the same subject matter".
 119. Witty, Paul A. and Sizemore, Robert A. "Studies in Listening, III: The Effectiveness of Visual and Auditory Presentations with Changes in Age and Grade Levels," Elementary English, XXXVI (February, 1959), 130-40.
Bases conclusions primarily on experimental evidence reported in 20 studies.

120. Woodcock, Richard W. "An Experimental Prognostic Test for Remedial Readers," Journal of Educational Psychology, XLIX (February, 1958), 23-27.

Describes the nature, development and evaluation of a test, composed entirely of pictures and non-letter symbols, for selecting remedial readers most likely to profit from special instruction in reading.

MEASUREMENT OF COLLEGE LEVEL READING COMPETENCE IN A CONTENT AREA

IRMA T. HALFTER and FRANCES M. DOUGLASS
De Paul University, Chicago

CONTENT AREA teachers, particularly on the high school and college levels, are being urged to teach the reading skills needed for mastery of their subjects. Such requests are often met with inquiries as to the nature of the reading skills required in a specific content area. Published reading tests seem to assess competence in general reading skills which may or may not be relevant to the particular area of concentration or may not reflect the peculiar emphases practiced in that field.

Problem

The goal set for this investigation was to construct a reading test with curricular validity appropriate for a collegiate school of business which, through its items or groups of items, measured the student's general competence in skills peculiar to the field of commerce. All items in the measuring instrument, then, would be derived from content materials in commerce and would attempt to reflect the patterns of collegiate schools of business generally. The test should also predict success in a collegiate school of business. The instrument, it was thought, should be a "power" analysis, every person having time to read and attempt all items. Items were to be of the multiple choice type, with five alternatives.

Background Research

To determine the nature of the curricular validity desired, the following investigations were undertaken:

1. Analysis of the kinds of reading skills that were implied by a) the questions of the mid-semester tests and final examinations, both objective and subjective, over a five-year period, of the faculty of the College of Commerce of De Paul University; b) the textual presentations in the fields of accounting, economics, finance, management, and marketing, and also social studies; c) the announced intentions and character of questions in workbooks

published to accompany standard texts.

2. A controlled project in the content fields of accounting and economics to determine whether the conclusions reached as to the nature of reading skills required were operative in the success or non-success of students in those fields.

3. Re-examination of the data of controlled analyses by the Commerce reading staff and faculty, collected over a five-year period, indicating that students with "developmental" reading difficulties as well as those with severe inadequacies in comprehension were primarily descriptive, existential readers.

Results of the Investigations

In the field of business and social studies at the college level, certain specific reading skills seem to be required. This conclusion was based on their use in the curricula, in the texts and in instruction. These skills tend, particularly in the specializations within the field of business, to be progressively employed in a text or at various levels of education. The measuring instrument should be composed therefore, of items not only in order of difficulty but, as far as possible, by groups of items, sectionally divided, to measure ascending levels of curricular difficulty. Each item should attempt to measure a single kind of skill. The general conclusion, issuing from the investigations, was: to be competent in reading in commerce involved being an analytical, essential reader.

Although examination of the literature measuring factors of reading comprehension did not seem to indicate that the thesis of specific skills would be supportable, it was determined, nevertheless, to assume ipso facto that they did exist and to construct a measuring instrument that would reflect the skills apparently required.

Procedure in Present Study

Construction of items: Although it may be dis-

* A footnote will be found at the end of the article.

analysis began with the computing of point biserial coefficients of correlation between responses on each item and the total score for each individual. A *t* ratio was then computed on the reliability of the point biserial, and all surviving items with a *t* value significant at the .001, .01, .02, and .05 level were tabulated for upper-lower group discrimination. All items were subjected to chi-square and Yates correction and *t* tests, and only those items found significant at the .05 or better level of confidence by both measures were retained. This reduced the items to 75 (one additional item was retained for student's heuristic purposes but is not scored). Table I contains selected illustrative calculations for the inclusion and exclusion of items.

The test was administered again to another random sample of 188 students at all collegiate academic levels. The item analysis, through computation of point biserial coefficients of correlation between responses on each item and the total score for each individual and the computation of the *t* ratio in the manner prescribed, showed all items receiving a *t* value significant at the .05 or better level of significance. These items also again survived the chi-square and Yates correction and *t* tests between the upper and lower quarter groups, with the same level of confidence by both measures previously found in the 75-item tests. Specifically, 44 items were at the .001 level, 18 at the .01 level, 4 items at the .02 level, and 9 at the .05 level on both measures. (See Table II).

Six months after the first administration of the test, the test was administered to 132 of the same standardization group (of 188). The test-retest coefficient of reliability was $r = +.93$, $\sigma_r = .016$; $M_1 = 49.28$; $M_2 = 49.53$. A coefficient of equivalence was taken, by splitting the test in half, through odd-even grouping of items of comparable difficulty. This split half reliability was: $r = +.81$, $\sigma_r = .031$; $M_1 = 23.79$; $M_2 = 23.76$; with Spearman-Brown correction: $r = +.90$. There is some evidence of "inter-judge" validity, since independent ratings of tested students by two trained reading counselors, one from De Paul University and one from the University of Illinois, correlated very highly with test standings: $r = +.86$, $\sigma_r = .026$.

A further analysis was made of the responses of the standardization group of 188 (the final form of 75 items) to see whether each section of the test was measuring a different level of skill. The *r* between Vocabulary and Sentence Meaning was .409; between Vocabulary and Paragraph Comprehension, .414; between Sentence Meaning and Comprehension was .551; between Sentence Meaning and Graphs, .441. The *r* between Paragraph Comprehension and Graphs was .528 (Table III): It would seem that the test does reflect, as it progresses from section to section, different and increasingly difficult levels of skills required for competent reading in commerce. To determine whether the test may be designated a "reading" test, a coefficient of correlation

between the Survey Section, Diagnostic Reading Test, Form D, and the Commerce Reading Comprehension Test, was obtained for this group of 188 students. The *r* was .747, $\sigma_r = .057$. It should be borne in mind that the Diagnostic Reading Test is a speed test.

Norms

Tentative norms (with an *N* of 90 for each of the ranks, freshman, sophomore, junior, and senior) showed all differences between means were in the expected direction and all were significant at the .01 level except that between the freshman and sophomore groups (.05 level). Regarding these groups, however, it was worthy of note that the freshmen showed a significantly higher (.01 level) degree of variability.

Predictor of Success

In an unpublished report, R. Duane Andrews, Office of Student Personnel and Guidance, University of Wyoming, has analyzed the role of the Commerce Reading Comprehension Test in predicting success in the College of Commerce and Industry at the University of Wyoming. High school grade averages, Ohio State University Psychological Test scores (Form 23), Commerce Reading Comprehension Test scores, and University grades were available for 78 of the 87 students to whom the battery of tests was administered. The test authors had stated, "The ability to read well is highly correlated not only with successful achievement in college but also with competence in many areas of business." Since no specific evidence of this correlation was available, the University of Wyoming undertook to study the relative validity of the Commerce Reading Comprehension Test, the Ohio State University Psychological Test, and high school grade averages as predictors of academic success.

The Ohio State University Psychological and Commerce Reading Comprehension Test both showed correlations with the criterion of college grades of .64, indicating about equal predictive value. The authors note that the Ohio State University Psychological Test and the Commerce Reading Comprehension Test correlate highly with each other (.61) as might be expected since each is heavily weighted with vocabulary and reading materials.

The combination of Ohio State University Psychological Test, high school grades, and the Commerce Reading Comprehension Test provides a multiple correlation of .77 with first semester grades in the College of Commerce and Industry, University of Wyoming. This *r* is significantly higher than those obtained using only the Ohio State University Psychological Examination and Commerce Reading Comprehension Test or the Ohio State Psychological Examination and high school grades, or the Commerce Reading Comprehension Test and high school grades. Among the author's

puted whether competence in vocabulary measures reading comprehension or intelligence, it was determined to measure vocabulary, through Section I of the Commerce Reading Comprehension Test. The selected word and the alternative answers were designed to measure the first and fundamental skill required in commerce and in the form practiced in commerce, defining structurally or functionally. No responses were to require choosing among "result of" judgments. Curricular demands were categorizing required reading skills, at their first level, by increment and in order of difficulty, into those concerned with definitions, distinctions, factors (as opposed to high school emphasis on characteristics), and direct cause-effect conclusions.

Items in Section II of the Commerce Reading Comprehension Test, headed "Sentence Meaning," were to measure chiefly the ability to obvert negatives into positives, interpret reversals of cause-effect relationships and the significance of qualifiers such as "unless" and "provided," as well as generalize through determining the meaning of a concept in context.

A separate grouping of these items was attempted because analytical data of the Commerce reading staff and the results of the controlled experiments in the content areas indicated that it was usually in the later stages of a text there occurred these negative propositions which were to be read affirmatively. Academic competence seemed to decrease when there was lack of understanding of this kind of expressed thinking and the implication, "It does not follow." Since the reading staff had found it was such a single sentence that confused the reader's total adequate grasp of the paragraph or section, it was determined to use the single sentence form in this section.

Section III of the Commerce Reading Comprehension Test, "Paragraph Comprehension," was to measure competence in reading the earliest and simplest, or primary, form of expression of thinking in a commerce textbook or course: the ability to recognize an author's deductive conclusion or to infer one when the facts were inductively presented, the latter being more difficult, apparently, for students to comprehend than the former. Such propositions would be concerned with judgments stating distinctions, factors, and cause-effect relationships.

This section would particularly concern itself with author's expressions of adversative or integrative modes. It would contain contrary propositions (almost uniformly identifiable in texts and scholarly thinking by "but" in the second proposition) and contradictory propositions (marked by "however" in the second proposition). The first occurrence of hypothetical and adversative propositions seemed to coincide with the beginning of decreased academic proficiency, in reading and teaching analyses, for approximately twenty per cent of developmental readers.

Since in evaluation of texts and teaching experience, the highest level of reading competence in-

volved the recognition of the dilemma or paradox and its solution by the author through a complex, qualified decision, items were included to measure the extent of competence in following that process of reasoning or completing the unexpressed implication. Although there might be some overlap of items with Section II of the test, it was nevertheless determined to include the items in a paragraph setting.

Only a few items were to measure competence in reading for details, since students generally had this skill or quickly acquired it, and because implicit in the curriculum of a collegiate school of business is the assumption that the student must always read for the undoubted significance of the facts, their single meaning, their rigorous interpretation.

In summary, then, items in Section III of the test, "Paragraph Comprehension," require the reader either to recognize a conclusion or mediate judgment deductively presented, or to connect an antecedent and consequent, or to draw inferences, or to imbricate partial or contrasting or contradictory conclusions, or to resolve alternative claims, or to distinguish "near right" from "quite right" implications.

Since competence in a school of business involves understanding the visual presentation of materials, items in Section IV of the test were designed to measure the skills at all levels through interpretation of graphically presented material. Use of arithmetic was kept to a minimum to make the task essentially an exercise in acquiring meaning through reading. The necessity for a reader to interpret graphs was a de facto judgment, out of teaching experience. The validity of the judgment of including a section on graphs was postponed until data were accumulated on the correlations between sub-scores and total scores.

Standardization, Reliability, and Validity

The first form of the test, consisting of 145 items, arranged in gross order of difficulty, was administered to a random sample of 353 students at all collegiate academic levels in the College of Commerce of De Paul University. An extensive item analysis began with the computing of point biserial coefficients of correlation between responses on each item and the total score for each individual. A *t* ratio was then computed on the reliability of the point biserial as an indication of reliability. All items receiving a *t* value significant at the .001, .01, .02, and .05 level were retained. An upper-lower quarter response frequency graph was constructed. The non-discriminating items were either revised, rewritten completely, or excluded in constructing the revised edition of the test.

The second form of the test, consisting of 122 items, arranged in gross order of difficulty, was administered to a random sample of 116 students at all collegiate academic levels. An extensive item

TABLE I

COMMERCE READING COMPREHENSION TEST--N = 116. ILLUSTRATIVE COMPUTATIONS
FOR SOME ITEMS INCLUDED AND EXCLUDED

Rank Order	Item Number	Frequency	Per Cent	Point Biserial	t Ratio	Level of Confidence	
1	1	186	.989	.066	.90	.40	*
44	4	18	.096	-.084	-1.15	.30	*
77	78	139	.739	.091	1.25	.30	*
84	92	121	.644	.113	1.55	.20	*
101	89	32	.170	.028	.38	.80	*
14	29	122	.678	.178	2.47	.02	**
20	38	94	.500	.327	4.72	.001	**
25	26	71	.378	.421	6.33	.001	**
73	87	155	.824	.222	3.11	.01	**
92	84	75	.399	.144	1.98	.05	**

* Items excluded from final test because level of confidence on Point Biserial and t ratio was not .05 or better.

** Items excluded from final test because these items did not survive chi-square and t tests in next analysis.

TABLE II

UPPER-LOWER GROUP DISCRIMINATION OF SIGNIFICANT ITEMS SURVIVING THE ABOVE
COMPUTING OF POINT BISERIALS AND t RATIOS--N = 188. ILLUSTRATIVE
COMPUTATIONS FOR SOME ITEMS INCLUDED AND EXCLUDED IN FINAL FORM OF TEST

Item Number	Frequency (Upper Group)	Frequency (Lower Group)	Chi ²	Level of Confidence	t Ratio	Level of Confidence
6	45	36	9.35	.01	2.809	.01
63	38	28	6.97	.01	2.31	.05
76	26	14	11.40	.001	2.60	.02
93	38	21	19.10	.001	3.93	.001
118	28	7	31.33	.001	5.08	.001

TABLE III
CORRELATIONS OF SUBTEST SCORES

Subtest I (Vocabulary)	Subtest II (Sentences)	Subtest III (Paragraph)	Subtest IV (Graphs)	Total
	$r = .409$ I-II $\sigma_r = .060$	$r = .414$ I-III $\sigma_r = .060$	$r = .369$ I-IV $\sigma_r = .063$	$r = .791$ I-T $\sigma_r = .027$
		$r = .551$ II-III $\sigma_r = .050$	$r = .441$ II-IV $\sigma_r = .059$	$r = .716$ II-T $\sigma_r = .036$
			$r = .528$ III-IV $\sigma_r = .053$	$r = .803$ III-T $\sigma_r = .026$
				$r = .717$ IV-T $\sigma_r = .055$

N = 188 $\Sigma I = 76$
R 1.34 = .449

$r \text{ II, III, IV-T} = .90$ R 1.23 = .466
 $\sigma_r = .014$

TABLE IV
CORRELATIONS BETWEEN THE AMERICAN INSTITUTE OF ACCOUNTANTS'
ORIENTATION TEST, FORM B, (A. I. A.) AND COMMERCE READING
COMPREHENSION TEST (C. R. C. T.)

Tests	Scores	N	M	σ	r
C R C T.	Total	213*	40.8	8.94	.8264
A. I. A.	Total	213*	50.2	20.19	
A. I. A.	Total	92**	53.45	18.78	.7877
C. R. C. T.	Total		42.75	7.94	
C. R. C. T.	Total	152***	39.95	7.512	.7221
A. I. A.	Total	152***	49.45	17.125	
C. R. C. T.	Total	152***	39.30	7.57	.6885
A. I. A.	L-Score	152***	27.85	12.186	

- * Incoming freshmen, Fall, 1956--no accounting courses.
 ** 92 of the original N of 213 who completed, January, 1958, three courses in accounting.
 *** Incoming freshmen, Fall, 1957--no accounting courses.

TABLE V

CORRELATIONS OF GRADE POINT AVERAGES AND PLACEMENT TEST TOTAL SCORES
(Freshman Class, 1952-1956, College of Commerce, De Paul University)

Test	Graduates				Total Freshman Group*			
	r	N	M	σ	r	N	M	σ
A. C. E.	.19	66	113.78	19.24	.34	166	97.69	22.92
Mathematics	.32	62	20.68	5.55	.45	154	19.95	5.80
English	.16	66	106.25	36.15	.25	166	95.85	40.05
Reading	.22	66	73.3	9.35	.40	166	70.00	12.15

* 166 including Graduates. Correlations are for students who stayed in college for varying lengths of time. Whereas 66 were graduated, 100 left some time before completion of the full college course. The r's include second, third, and fourth year grades.

A. C. E. --American Psychological Examination, 1949 Edition.
Mathematics--Cooperative Mathematics Pre-Test for College Students, Form Y.
English--Iowa Placement Examinations, English Training, Form M.
Reading--Survey Section, Diagnostic Reading Tests, Form D.

TABLE VI

TEST INTERCORRELATIONS--TOTAL FRESHMAN GROUP

Criterion*	A. C. E.	Reading	Math.	English
C.	.34	.40	.45	.25
1.		.68	.33	.49
2.			.30	.67
3.				.53

* Grade Point Average. Correlations are for students who stayed in college for varying lengths of time. Whereas 66 were graduated, 100 left some time before completion of the full college course. The r's include second, third, and fourth year grades.

conclusions is the judgment that the use of the Commerce Reading Comprehension Test reduces the relative effectiveness of high school grades as a predictor of college success, while improving the general predictiveness of the battery.

The college averages were computed for all freshmen in their first year of study, 1956-1957 (the first year the Commerce Reading Comprehension Test was administered to all incoming freshmen in the College of Commerce of De Paul University). The multiple r for the Commerce Reading Comprehension Test and the Cooperative Pre-Test for Mathematics, Form Y, and the criterion was .56.

Other Correlations

In September, 1956, as part of a battery of tests administered in the College of Commerce of De Paul University, 343 freshmen took the Cooperative School and College Ability Tests (SCAT), Form 1-C, and the Commerce Reading Comprehension Test (CRCT). The correlation between the total score of the SCAT test and the CRCT total score was: $r = .74$ (σ SCAT = 11.787; σ CRCT = 9.339).

This correlation was to be expected, since the verbal skill required in the SCAT test is dependent in part on competence in reading. Rather more significant, or perhaps equally interesting, is the nature of the items in Part I of the SCAT Test. Their understanding requires ability to recognize and resolve antecedent-consequent, hypothetical or conditional, contrary, and contradictory judgments.

A correlation was also obtained between the American Institute of Accountants' Orientation Test, Form B (A.I.A.), total score, and the Commerce Reading Comprehension Test (CRCT), total score, for an N of 213 incoming freshmen, September, 1956. The $r = .8264$ ($\sigma r = .02$) (σ A.I.A. Test, = 20.19; σ CRCT = 8.94). Since this correlation is surprisingly high and may be a reflection of coincidence, it was decided to secure a correlation of a selected N of 92 (out of the original N of 213) who had completed three semesters of accounting by the end of January, 1958. The $r = .7877$ (σ A.I.A. Test, = 18.78; σ CRCT = 7.94).

A correlation was also obtained between the same A.I.A. Test, total score, and the CRCT total score, for the 152 incoming freshmen administered both tests, September, 1957. The $r = .7221$ (σ A.I.A. Test, = 17.125; σ CRCT = 7.512). The correlation between the same A.I.A. Test, L-score, and the CRCT total score, for the same N of 152, September, 1957, was: $r = .6885$ (σ A.I.A. Test, = 12.186; σ CRCT = 7.57). (Table IV)

Post-Research

In a longitudinal study of the class of 1952-1956 of the intellectual and certain socio-economic factors operative, perhaps, in the academic performance of students in the College of Commerce of

De Paul University, correlations between grade point averages and test scores as well as test inter-correlations and multiple r 's for all the tests used were established (Table V, VI). These were affected by the heterogeneity of the De Paul Commerce group, with varying lengths of residence; the clustering of the grade averages, for the graduating group would include no C-minus students and the general group of freshmen would include those retrained in reading, English, and mathematics; and the nature of the curricula. (See Table VI).

Bearing in mind these limitations upon interpretation of the correlations, it was nevertheless concluded that the result of the curricular demands on training in a specialized field may be reflected. An r of .19 (ACE) and an r of .22 (Diagnostic Reading) for the graduating group as against an r of .34 (ACE) and an r of .40 (Diagnostic Reading) for the total freshman group may be indicative, in part, of the effect of curricular demands. College of Commerce students have many courses in common with Liberal Arts students, but as they progress from freshman to senior year, there is an increasing disparity in course content with possible emphases on certain other kinds of reading skills than those stressed in the (then) entrance testing. There is, of course, the possibility that too high an r between scores obtained at entrance and final grade averages for the graduating group may indicate a lack of certain types of growth and development.

Results of the Present Study

The results are sufficiently encouraging to continue holding to the theses that there is emphasis on certain kinds of reading skills in a collegiate school of business and that these skills seem to involve more than what is generally regarded as comprehension in reading. Further inquiry will isolate other or additional skills than those measured, and a better measuring instrument should be devisable. Sufficient experience has not been had to know whether the items or the Commerce Reading Comprehension Test as a whole gives a sufficiently refined judgment. A curricular study should be made to determine whether these or other levels of skills are common to other areas of concentration. Accumulating evidence at De Paul University, College of Commerce, would indicate that achievement tests, that is a reading test of the power type that measures competence in the reading skills outlined, plus a mathematics test, may be a better predictor of success in college than ability tests.

FOOTNOTE

* Acknowledgment is due Dr. Raymond J. McCall, former Director of Testing at De Paul University, now at Marquette University, and who is co-editor of the De Paul Commerce Reading Comprehension Test; and Dr. R. Duane Andrews, Office of Student personnel and Guidance, University of Wyoming,

for his unpublished analysis of the role of the Commerce Reading Comprehension Test and other tests

in predicting success in the College of Commerce and Industry at the University of Wyoming.

CHANGES IN AFFECT ATTRIBUTABLE TO INSTRUCTION IN READING IMPROVEMENT AT THE COLLEGE LEVEL

MERYL E. ENGLANDER
Sacramento State College

EFFECTIVE READING is the obvious goal of reading improvement programs. Smith (6, p. 5) suggests that reading improvement courses give instruction on how to read. In such a program, weaknesses would be diagnosed and help given in improving speed and comprehension, budgeting time for study, developing ability to concentrate, building vocabulary, finding and organizing information, and attacking and solving problems. Smith and Wood (5, p. 158) found in a rigorous study that students who participated in such a reading improvement program show significant gains in reading performance, that such gains are maintained and possibly increased after a lapse of time, and that the gains are reflected in significant superiority in academic status.

One other factor in effective reading deserves attention. The attitude which the individual has toward himself as a reader and toward reading as an activity is important. Bond and Tinker (1, p. 395) point out that the acquisition of reading ability is of little value to a person unless he puts it to use in reading various types of materials spontaneously and voluntarily. Strang *et al.*, (7, p. 6) indicate that they find: "Many poor readers are held back by fear... A vague anxiety may give rise to feelings of inadequacy, helplessness, and hopelessness. Their idea of themselves is permeated with lack of self-confidence and self-esteem." Fernald (2, pp. 7-18) has emphasized the need for self-confidence and feelings of success as a necessary component of reading improvement.

Changes of attitude in persons who complete a course in reading improvement are detectable. Students indicate effective changes through such statements as: "I never cared much for reading before, but now I find that it is stimulating" or "I never read a book for pleasure before taking the course in reading improvement, but I read three books last summer just for fun" or "That course in reading sure helped me; I'm now getting much more out of school."

Such statements as these are gratifying to the instructor and suggest a favorable change in attitude

toward reading but are they valid measures of change in feeling? Are the changes significant? In which areas specifically do students change their attitudes as a consequence of instruction in effective reading?

This paper attempts to answer these questions through an exploratory study of changes in attitude toward reading and allied subjects in students who participate in a reading improvement program.

The Instrument

Appropriate instruments for the measurement of attitudes toward reading are not available. Therefore, it was necessary to devise a suitable measure to evaluate this quality.

The instrument for this study was developed in the following way: An open ended questionnaire was administered to one hundred students who had enrolled in a reading improvement course. The questionnaire consisted of twenty words and phrases which had been selected on an *a priori* basis for their ability to stimulate responses which would disclose the subject's attitude toward reading and related academic matters. Such words or phrases as the following were presented: Words-, Examinations are-, and I like to read-.

The responses from this open ended questionnaire were used to construct a forced-choice scale. The previously mentioned words and phrases were used as leads and selected responses were used as alternatives. The alternatives were graduated to represent increments of attitudes from positive to negative. Each subject was asked to select the one response to each item which was most representative of his attitude. The scale is presented, starting at the top of the next page.

No attempt was made to extend the scale limits to include every extreme attitude. Very possibly attitudes more extreme than the alternatives permit were felt by some of the subjects. This limitation of expression would tend to limit the amount of change which could be measured.

Likewise, no attempt has been made to calculate

INSTRUMENT FOR MEASURING STUDENT'S ATTITUDE

Indicate which of the respective choices expresses your feelings:

1. I expect Reading Improvement to:
 1. vastly improve my reading speed, comprehension and vocabulary.
 2. help me be a better reader.
 3. be of some use.
 4. give me two units of credit without too much effort.
 5. be of little value; I'm taking it under protest.

2. My plans for future education are:
 1. to do graduate work after I have had my Bahcelor's Degree.
 2. without question to finish college.
 3. to finish college if all goes as expected.
 4. such that I probably will not finish college.
 5. not to continue for a college degree.

3. The thing that I dislike about school is:
 1. not having the time to really explore subjects.
 2. having to take courses which are not challenging.
 3. having to take courses which are of no value.
 4. not being able to understand the work.
 5. the studying.

4. Words:
 1. are intriguing.
 2. are used to express ideas.
 3. are necessary.
 4. make reading difficult for me.
 5. are my downfall.

5. The thing which I like best about college is:
 1. the opportunity to learn.
 2. discussing intellectual problems with other students.
 3. it gives one a chance to obtain a better vocation.
 4. the opportunity to participate in extracurricular activities.
 5. the coffee breaks.

6. College is:
 1. interesting and a challenge to me.
 2. a good means of learning and getting ahead.
 3. okay for the time being.
 4. really quite a bore.
 5. quite unsatisfactory.

7. Intellectually speaking:
 1. I'm superior to most college students.
 2. I'm above the average college student.
 3. I'm about of average college intelligence.
 4. I'm a little less intelligent than most.
 5. I'm inferior to most college students.

8. Examinations:
 1. give student an opportunity to show what they know.
 2. are challenging.
 3. are a necessary evil in school.
 4. make me feel uneasy.
 5. are terrifying.

9. I prefer a course:
 1. that affords an opportunity for independent study.
 2. in which I can learn new material.
 3. that is interesting to me.
 4. that will be of practical use to me.
 5. that doesn't require too much reading.
10. My favorite pasttime is:
 1. reading.
 2. intellectual discussions.
 3. playing and listening to music.
 4. athletics.
 5. watching movies or television.
11. As a reader I am:
 1. good.
 2. better than the average college student.
 3. about as efficient as other college students.
 4. poorer than the average college student.
 5. poor.
12. The way I feel about reading is:
 1. I would rather read a book than do most anything else.
 2. I like to read whenever I have time.
 3. it is okay as a means of learning.
 4. I don't care much for it.
 5. I really dislike to read.
13. Studying:
 1. is easy; I'm amazed at how easy college has been.
 2. is fairly natural to me.
 3. doesn't give me much trouble.
 4. is hard, but I get it.
 5. is very difficult for me.
14. Time:
 1. I have plenty of time to do all of the things I wish to do.
 2. I do most of the things I wish to do.
 3. I usually have time to do the things I have to do.
 4. I usually have a need for more time.
 5. I never have enough time.
15. I prefer to read about:
 1. intellectual deductions.
 2. historic events.
 3. athletics.
 4. adventure.
 5. romance.
16. Most of my pleasure reading is:
 1. nonfiction.
 2. fiction.
 3. Reader's Digest and comparable magazines.
 4. Newspapers.
 5. Comics.

the scale values by the use of the equal-appearing intervals technique. Alternatives were placed in a hierarchy on an arbitrary basis. The instrument was scored by awarding the alternative responses a value of 5, 4, 3, 2, or 1 respectively as a range from the most positive to the most negative attitude.

The reliability of the instrument was determined by the test-retest technique with 30 subjects over an interim of four weeks. The Pearson correlation coefficient was .98.

The Findings

The instrument was used to measure change in attitude for persons who complete a course in reading improvement. The course as presented is typical in that major emphasis is placed upon effective reading. Topics such as regression, concentration, use of time, and vocabulary are discussed. Practice in effective reading constitutes a major portion of the student's effort. No direct attempt is made to change attitudes *per se*. The class meets twice weekly for a semester and carries two units of credit. The grade in the course is a function of improvement in reading rate and comprehension, effort in workbooks, and attendance.

The instrument was administered to 46 subjects on the first and last meetings of the class. The subjects were informed that the expression of attitudes would in no way influence their standing in the course. The results are presented in Table I.

The difference between means on the precourse and the postcourse administrations showed significant differences in several areas. Differences at the five per cent level of confidence are discernable for: self-concept in reading, attitudes toward reading, self-concept in studying, and attitudes toward school. The latter difference was in favor of the precourse mean. The meaning of these differences will be discussed in Conclusions.

Virtually no changes were apparent in several aspects explored by the instrument. For example, no changes were manifested in the students' attitude toward future education, academicness, intellectualism, or use of time.

The instrument was also administered to thirty students in an elementary public speaking course to ascertain if the above findings were a function of factors other than participation in a reading improvement program. The instrument was administered at the beginning and again at the close of the semester. As indicated in Table I, there were no significant differences in attitude evident between the precourse and postcourse measures for this control group.

Table I also presents data comparing the attitudes of those persons who feel reading inadequacies to the extent that they elect to participate in a reading improvement program and the random college student as found in an elementary speech course. These two groups differ in plans for future education at the one per cent level of confidence.

They also differ at the five per cent level of confidence in their attitude toward reading, their reading preferences, and their self-concept in reading and studying.

The significance of these findings will be discussed in the Conclusions.

The Conclusions

Student reactions seem to indicate that in addition to changes in reading skill, vocabulary, and study effectiveness certain changes in attitude are associated with participation in a reading improvement program. This paper is an exploration of a measure of this change in attitude.

Attitudes toward words, reading, personal ability in studying, and reading as measured in this study were improved for those in the reading improvement program. Attitudes toward reading and the self as a reader which Bond and Tinker (1, p. 395), Strang *et al.*, (7, p. 6) and Fernald (2, pp. 7-18) assert to be so important would seem to be changeable through instruction in reading improvement. This improved attitude in conjunction with improved reading skill should help the student to be a more effective reader.

This change in attitude is especially significant when it is noted that a like change was not evident in those persons who participated in a speech course and not in a reading improvement program.

The difference between those persons in a reading improvement program and those persons in the speech course are pertinent. Presumably, those persons in a course for reading improvement have elected or been counseled to the course because of reading inadequacies. Strang *et al.*, (7, p. 6) suggest that such persons have detrimental attitudes toward reading, themselves as readers, and other associated factors. The data found in this study support this contention. At the beginning of the semester, those persons in the reading improvement course had less desirable attitudes toward reading, reading preferences, and their self-concept as readers and students than did those subjects from the speech course. Interestingly, these are the attitudes most responsive to change in the reading improvement program as shown by the data in this study.

The apparent undesirable change in attitude toward school is difficult to explain as a consequence of the course in reading improvement. Rather, this phenomenon which is a general attitude involving all of college experience is more likely explained by noting that the precourse attitude was ascertained at the beginning of the semester when ambitions, resolutions, and hopes are high; while the postcourse attitude was ascertained at the end of a semester of reality. A further consideration is that though there is a statistically significant change, the postcourse attitude toward school is still relatively positive and equivalent to that held by the students in the speech course.

TABLE I
MEASURES OF STUDENT ATTITUDE TOWARD READING AND RELATED SUBJECTS

Item	I. Subjects from Reading Improvement Course, N=46			II. Subjects from Speech Course, N=30			III. Comparison of Precourse scores for Subjects I and II t score
	Precourse Mean	Postcourse Mean	t score	Precourse Mean	Postcourse Mean	t score	
1. Expectations from course in reading improvement	4.49	4.51	.08				
2. Plans for future education	3.40	3.29	.52	3.84	3.90	.27	3.02**
3. Academic nature of undesirable college features	3.29	3.73	1.45	3.84	4.07	.76	1.71
4. Words	3.47	3.95	1.96	3.67	3.74	.23	.81
5. Academic nature of desirable college features	4.33	4.49	.82	4.44	4.67	1.07	.48
6. College	4.78	4.51	2.06*	4.46	4.50	.18	1.99
7. Self-concept (Intelligence)	2.84	2.93	.18	3.06	3.17	.72	1.63
8. Examinations	3.80	3.79	.02	3.33	3.40	.20	1.70
9. Academic nature of desirable courses	2.87	3.16	1.53	3.03	3.23	.75	.72
10. Academic nature of use of leisure time	2.96	3.09	.51	3.50	3.60	.28	1.83
11. Self-concept (Reading ability)	2.42	2.76	2.02*	2.90	3.17	1.37	2.61*
12. Reading	3.35	3.73	2.12*	3.81	4.04	1.02	2.48*
13. Self-concept (student)	2.04	2.42	2.20*	2.56	2.76	.12	2.60*
14. Available time	2.82	2.84	.06	2.80	2.90	.33	.36
15. Academic nature of reading preferences	2.80	2.89	.46	3.40	3.66	.59	2.16*
16. Intellectual taste of pleasure reading	3.56	3.58	.10	3.77	3.74	.11	.69

* Significant at the 5% level of confidence.

** Significant at the 1% level of confidence.

No statistically significant changes are noted in attitudes pertaining to plans for future education, the academic nature of preferred courses, or the attractiveness of academic learning. Perhaps the experiences offered in a single course in which the emphasis is upon techniques is too limited to expect significant changes in these areas. Jacobs (3) in a thorough review of studies of attitude and value change in college indicates that significant changes in academic attitude are rare and then result only through a complicated psychological experience.

The absence of changes in attitude toward reading preferences, course examinations, and the use of leisure time is disappointing; particularly since these topics were discussed as part of the content of the course. Perhaps Smith's and Wood's findings (5, p. 438) that improvement in reading effectiveness is a function of personality components is an important factor in this issue.

Though not conclusive, this study does suggest that some affective changes are possible in a reading improvement program at the college level. It behooves us to explore this factor of reading effectiveness more thoroughly.

REFERENCES

1. Bond, G. C. and Tinker, M. A., Reading Difficulties: Their Diagnosis and Correction. New York: Appleton-Century-Crofts, 1957.
2. Fernald, Grace M., Remedial Techniques in Basic School Subjects. New York: McGraw-Hill, 1943.
3. Jacobs, P. E., Changing Values in College. New York: Harper, 1957.
4. Smith, D. E. P. and Wood, R. L., "Noncognitive Personality Dimensions as Indices of Reading Improvement at the College Level." American Psychologist, 1953, 8, 438.
5. Smith, D. E. P. and Wood, R. L., "Reading Improvement and College Grades: A Follow Up." Journal of Educational Psychology, 1955, 48, 151-159.
6. Smith, Nila B. (Gen. Ed.) Effective Reading for College Students, by Carter, H. L. J. and McGinnis, Dorothy J., New York: Dryden, 1957.
7. Strang, Ruth; McCullough, C. M. and Traxler, A. E., Problems in the Improvement of Reading. New York: McGraw-Hill, 1955.

SURVEY OF THE RESEARCH RELATED TO THE READING ABILITY OF THE GIFTED

SISTER JOSEPHINA, C. J. S.
Boston College
Chestnut Hill, Massachusetts

Studies of reading interests and abilities of the gifted pupil appear frequently in the research literature. For the most part these studies fall into the descriptive category, presenting the types of program best suited to the needs of the gifted; delineating the characteristics of their teachers; and explaining school plans in current operation. Few studies offer objective data substantiated by statistical treatment related to reading of the above average pupil.

The present problem summarizes some of the major findings in reading performance as related to the gifted pupil especially in the elementary grades.

The first exhaustive study on the gifted, begun in 1925 and completed in 1947, included research on reading habits of the gifted pupils who, according to Terman, learn to read easily, read more and better books than the average child and thus educate themselves (16:56). Terman, as part of his investigation of 1000 gifted children, found that they did not reach but approximated their reading capacity, showing better performance in reading than in arithmetic (17: 289-293).

The significant problem of the failure of the school to provide sufficient, systemic guidance for the gifted, especially in reading, is treated by Strang who indirectly castigates the elementary school for the ineffective teaching of reading and study habits resulting in the academic failure of high school and college students (14:24). She likewise states that the gifted pupil may be a retarded reader because he may be achieving below his potential (14:23). In another report Strang indicates that "know-how in reading" is not native and to trust that the able student will learn to read by trial and error appears quite unjustified (14:7).

That the bright pupil is retarded in reading is further described by McCullough, Strang, and Traxler (10:357) who stress that planned and guided work be given the bright pupil in the reading skills. These may be summarized as understanding the use of various reading aids; pictures, charts, diagrams; use of indexes; glossaries; sectional and marginal headings (10:130). The text is very worthwhile in the over-all presentation of the reading abilities needed

by the more able student. Thus, no teacher can assume that the above-average pupil will master independently the skills needed for research and higher study which will be his lot because of his high mental ability.

Corroborating Strang on the question of failure as related to the reading of the gifted are Monroe who found between 10 and 15 per cent of the elementary school population reading below that which is expected from their mental age (12:105-113) and McKee who stressed grades four, five, and six as strategic areas for promoting wide interest in reading and that pupils in these grades are not acquiring adequate control of the fundamentals of reading (11:348). Similarly, Gates in a survey of 12,000 pupils of whom were above average ability in grades 4, 5, and 6 found that approximately 30 per cent had not reached fourth grade level in reading (6:109-123).

To prevent the failure of the gifted pupil, Gray recommended that the achievement and needs of the superior pupil be checked regularly so that continuous growth in the fundamental skills be in harmony with their mental ability (7:755). Gray likewise cautioned that a major problem faced by the school is to devise educational programs which will make full use of the potential of the bright pupil (7:738).

In a well executed testing program of 2,918 pupils whose mental ages ranged from 6.10 to 17.4, Thomas found reading grade levels from 2.8 to 11.3 on the Stanford Achievement Test, Intermediate, Form DM with one out of every eight pupils failing. Approximately 65.5 per cent were working up to mental capacity (18:30). In the same article attention is focused on the 13.4 per cent, or 392 pupils with high mental ability, who are retarded more than one year below the expected level (18:32).

Frequently, the emotional factor as related to the reading ability of the superior pupil has been studied. In a large scale examination of two thousand pupils in grades four through seven in Cleveland, Sister M. Vera found no significant relation between personality factors and reading achievement. She stated that pupils in the lowest quartile range of the various personality components were not more retarded in reading than those ranking in the third quartile (19:161).

A similar study by Blackham reported opposite findings in that emotional disturbances and reading difficulties are related. He stated that under-achievers in reading gave slightly greater indications of anxiety arising from the frustrations of affectional satisfactions than were displayed by the over-achievers. The top centile group in reading showed a more mature and ready control over emotional impact (3:135). In Sister M. Vera's study the expected age obtained from Horn's study (8) was used while in Blackham a reading achievement index (Reading Age \div MA) was the prediction formula.

Baker summarized the findings related to personality and reading and reported that the majority of the research studies show that reading success is definitely associated with the wholesome adjustment of the child (2:13). In summary, because of his superior talent, the gifted show less emotional impairment where studied with reading disability. It may well be that the superior pupil appears less often than does his slow-learning peer in a remedial reading situation.

Analyzing the characteristics of the gifted child, Gray presented data that these pupils possessing certain well-defined characteristics which make it possible for them to excel quantitatively and qualitatively (7:737). However, Monroe and Backus found no clear cut factors which occur only in poor readers and never in good readers (13:12). Because the gifted is first a child with all the personality characteristics he will react as do all children whether slow, average, or bright. The main differentiable of the traits of the gifted is in the amount-quantitatively speaking—with certain interests and well-defined traits as learning by complex association, application of abstract thinking, and their strong ability to be self-critical. These are mainly applicable to the mental profile of the gifted.

Witty emphasized the role of teachers in improving reading instruction for the gifted and stated that the field of reading offers a great opportunity for enriching the background and experience of the gifted (20:147). The verbal ability of the superior pupil is his strength, mentally speaking, and, therefore, reading is an avenue offering full scope toward challenging, stimulating and satisfying his mental cravings.

Bliesmer found that bright children were significantly superior to dull children of comparable mental ages in total reading comprehension, some specific reading abilities, memory for factual details, perception of relationships, listening comprehension. However, no significant differences occurred when bright were compared with dull children of comparable mental age in reading rate, word recognition, and word meaning (4:329).

To develop reading skills of critical thinking, power to generalize, and to use reading as a resource tool in daily living were some objectives set forth for teachers in the middle and upper grades (1:164). Bright pupils must be taught reading skills

developmentally, and practice and drill where needed must be applied at certain learning points if the fullest returns in accomplishment are to result for the superior pupil. Jay is of the opinion that the teacher "seldom" guides the extra reading of the gifted student who is frequently allowed to read at random for entertainment only (9:37).

In analyzing the reading performances of one hundred gifted and one hundred average ability students in grades 5 and 6, Concannon applied the prediction formula as given by Horn (8) and estimated the expected reading age of both groups. In every comparison the gifted differed significantly from the average in reading vocabulary and comprehension as evidenced by the chi-square results (5). Yet, when the actual reading level was compared with the expected reading level the gifted evinced more retardation than the average when analyzed by the rigorous statistical "F" ratio (5:124).

The summary presented a brief review of the research relating to the reading of the gifted. The large segment of the literature is marked by an absence of statistical treatment dealing with the problem. Clearly defined and objectively presented studies will contribute to a more valid understanding of the problem.

REFERENCES

1. Adams, Elsie W. "Guiding Reading Activities of Superior Learners in the Middle and Upper Grades," Classroom Techniques in Improving Reading. Chicago: University of Chicago Press, 1949, pp. 162-166.
2. Baker, Norma. Characteristics of Superior and Retarded Readers in the Intermediate Grades of the Elementary School (unpublished Ph.D. dissertation, Department of Education, University of California, Berkeley, 1951).
3. Blackham, Garth J. A Clinical Study of the Personality Structures and Adjustments of Pupils Under-Achieving and Over-Achieving in Reading (unpublished Ph.D. dissertation, Department of Education, Cornell University, Ithaca, 1954).
4. Bliesmer, Emery. "Reading Abilities of Bright, and Dull Children of Comparable Mental Ages," Journal of Educational Psychology, XLV (October, 1954), 321-331.
5. Concannon, Sister Josephina, CSJ. Relationship of Actual Achievement to Potential Achievement in Reading and Study Skills of Gifted and Average Pupils (unpublished Doctoral dissertation, Boston College, 1957).
6. Gates, Arthur I. "The Psychological Basis of Remedial Reading," Educational Record Supplement, XVII (October, 1936), 109.
7. Gray, William S. "Education of the Gifted Child with Special Reference to Reading," Elementary School Journal, XLII (June, 1942), 737.
8. Horn, A. McA. Uneven Distribution of the

SISTER JOSEPHINA

- Effects of Specific Factors, California: University of Southern California Press, 1941.
9. Jay, Edith. "Types of School and Class Organization: Their Advantages and Limitations," Promoting Maximal Growth Among Able Learners, Chicago: University of Chicago Press, 1954.
 10. McCullough, Constance M., et al. Problems in the Improvement of Reading. 2d ed. revised. New York: McGraw Hill Co., 1955.
 11. McKee, Paul. The Teaching of Reading. Boston: Houghton Mifflin Co., 1934.
 12. Monroe, Marion. "Diagnostic and Remedial Procedures in Reading," Educational Record Supplement, XX (January, 1938), 105-113.
 13. Monroe, Marion, and Backus, B. Remedial Reading. Boston: Houghton Mifflin Co., 1937.
 14. Strang, Ruth. "Gifted Children Need Help in Reading, Too," The Reading Teacher, VI (January, 1953), 23-27.
 15. Strang, Ruth. "Basic Issues and Problems in Reading Instruction for Capable Students" Promoting Maximal Reading Growth Among Able Learners. Chicago: University of Chicago Press, 1954, 6-10.
 16. Terman, Lewis et al. Genetic Studies of Genius. 4 vols. Stanford, California: Stanford University Press, 1925-1947.
 17. Terman, Lewis et al. The Intelligence of School Children. Boston: Houghton Mifflin Co., 1919.
 18. Thomas, G. I. "A Study of Reading Achievement in Terms of Mental Ability." Elementary School Journal, XLVII (September, 1946), 28-33.
 19. Vera, Sister Mary. "A Critical Study of Certain Personality Factors as Determining Elements in a Remedial Reading Program," Catholic Educational Review, XL (March, 1942), 145-161.
 20. Witty, Paul. "What Can Teachers Do to Improve Reading Instruction for the Gifted Pupil?" Education LXXV (November, 1954), 145-149.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 53

MARCH, 1960

NUMBER 7

CONTENTS

Evaluations of Study-Skills Courses: A Review Doris R. Entwisle	243
Authoritarian Trends in Personality as Related to Attitudinal and Behavioral Traits of Stu- dent Teachers—Joseph A. Del Popolo.....	252
Reading Skill as a Predictor of College Achieve- ment—Malcolm H. Robertson and Mildred M. Harrison	258
Extraversion, Neuroticism, and Student Achievement in Introductory Psychology— A. W. Bendig	263
The Effect of Sonic Emotion on the Immediate Recall of Nonsense Syllables—Richard H. Bloomer and Sorrell E. Chesin.....	268
Principles of Child Growth and Development— Charlene D. Palmer.....	273
Placement Bureau Data and Teaching Success— Maurice F. Foss.....	276
Are Present Reading Tests Valid for Both Girls and Boys?—Walter J. Pauk.....	279

DEMBAR PUBLICATIONS, INC.
MADISON, WISCONSIN

EVALUATIONS OF STUDY-SKILLS COURSES: A REVIEW

DORIS R. ENTWISLE *
Towson, Maryland **

Introduction

There has been an increased tendency in the past ten years for colleges and universities to offer courses in study techniques. Some of these courses are devoted exclusively to problems connected with reading, but others define study problems very broadly, and include even extensive individual counseling as part of a course. Blake (2), from results of a survey he conducted in 1953, reports that over 90 per cent of colleges in the United States offer some kind of a study skills course, at least to selected groups of students, and 10 percent require such a course of all Freshmen. It is the purpose of this review to examine reports of evaluations of study skills courses to see how effective the courses actually are.

The notion that improved study habits, leading in turn to increased academic effectiveness, might result from study-skills courses is intuitively appealing, and one does not have to look far in the educational literature to find some rationale for offering these courses. Experiments on transfer of training show that some very general kinds of study skills, such as reasoning, reflective thinking, and ability to memorize, can be taught and effectively transferred to new situations. As early as 1945 it was becoming clear that remedial reading courses at the college level were probably helpful.

The issue of what specific techniques should be included in a study-skills course is unsettled because of a dearth of clear-cut research findings. Independent experimental investigations of even relatively simple problems, such as the optimum method of reading for study purposes, have led to different recommendations. As a consequence, for many of the topics they treat, the numerous texts and manuals on study techniques express "expert opinion" rather than results of empirical findings.

There has been considerable work of late on study habits inventories, and it may turn out that further work with these will yield a more rational solution

to what study courses should include. Although these inventories are still in a preliminary stage of development, there is considerable evidence that these tests measure attributes which are important determiners of achievement (3, 4, 16, 17, 22). Carter (4-8) has worked extensively with study habits inventories, and his factor analyses suggest that the tests tap four principal variables: (1) morale or self-confidence, (2) scholarly drive and values, (3) study mechanics, and (4) tendency to plan for getting work done. Initially, Carter felt that morale and scholarly drive were most important, but recent work by both Carter (9) and Chahbazi (10) tends to show that study mechanics is more important than was formerly believed. Since study mechanics was harder to measure than the attitudinal variables, its importance was underestimated.

Previous work, then, indicates that courses are probably beneficial, that study methods contribute to achievement, but can offer little help on which methods are optimal. It has therefore been incumbent upon those administering study skills courses to carry out evaluations of the courses.

Evaluations of Study-Skills Courses

A total of 22 evaluations, which include detailed data, were found in the literature. Table I summarizes the results of these evaluations. Before we can interpret the results, it is necessary to consider first a few aspects of the design of these studies.

The Criterion - The criterion used to determine whether study-skills courses were effective differed somewhat among the evaluations, although all (except for two rather specialized studies) included a measure of overall scholastic average. It seems mandatory that there be an adequate criterion of the departure of the overall average of an experimental group from a baseline provided by a suitable control group. (More will be said below about what comprises a suitable control group.) If grades are available for several semesters in addition to the

* The writer is indebted to J. M. Stephens and D. G. Danskin for critical readings of the manuscript
** Present Address: 802 Shaw Court.

TABLE I
SUMMARY OF EVALUATIONS OF STUDY-SKILLS COURSES

Study	Kind of Student	Criterion Used	Baseline Used	Methods Used to Correct for Bias	Performance in Excess of Baseline	
					Immediate	Follow-up
Studies Composed of Volunteer Students						
(1)* Barbe	Freshmen through senior law students	Reading scores; - grade-point average	Equal number of students unable to schedule course	Both groups given initial reading test	Significant increase in reading rate and grade-point average (.05 level)	Six months later improvement maintained
(11) Charles	Mostly Freshmen	Grade-point average	Group not taking course	Groups matched on: sex, previous grades, ACE	65% trained group better average than control, not statistically significant	Maintained gain second semester
(20)* McDonald	College students	Grade-point average	Group <u>desiring</u> to enroll in course	Covariance analysis to allow for differences: age, ability, initial reading rate	Significant increase in grade-point average; significant fewer dropouts	Cumulative 2- and 3- semester averages significantly increased
(21) McGinnis	College students	Point-hour ratios; reading scores	Group not taking course; own previous reading scores	Groups matched on: sex, college year, scholastic average, reading	Significant increase point-hour ratios (.01 level); significant increase reading scores	--
(26) Ranson	College students	Grade-point average; reading scores	Group not taking course; own previous average	Groups matched on: sex, age, division of university grade-point average; not on reading	Significant difference in grade-point average and reading scores	Difference between groups increases later
(27) F. Robinson	College students	Skill tests, grades, social adjustment, behavior rating	Own previous performance	"matched control" --not specified	"definite improvement" in 4 criteria listed	--

(24)* H. A. Robin- son	Freshmen	Total grade aver- age	Group desiring to enroll in course	Matched on: read- ing tests, intelli- gence, aptitude, secondary achieve- ment	Difference (.10 level) in grade average	--
(29) Sharp	College students	Grade-point aver- age	Group not taking course; own pre- vious grades	Matched on: pre- vious marks, IQ, reading scores, and other character- istics	Average gain of 9.07 grade points over own previous record, 11.37 points over control group	--
(32) Simpson	Freshmen	Grade-point aver- age	Group not taking course	Matched on ability	Trained group average is one-quarter letter grade better than control	--
(33)* Smith and Wood	Freshmen	Grade-point aver- age; reading scores; drop-outs	One group desir- ing to enroll; 2nd group representa- tive sample of class	Matched on ACE scores	Significant increase grade-point average and reading scores over both control groups; significant fewer drop- outs	Maintained gains 14 months later
(38) Witten- born	Freshmen	Predicted expect- ed scores	Expected scores; poor readers not attending	Matched on: poor reading scores, aptitude (?)	Exceeded expected scores, while class as whole failed to attain; significant gains in reading scores	--
(31) Sheldon	Medical students	Reading scores	Own previous scores	--	Median growth of 22 centile points in read- ing scores	--

(Continued)

TABLE I (Continued)

Study	Kind of Student	Criterion Used	Baseline Used	Methods Used to Correct for Bias	Performance in Excess of Baseline	
					Immediate	Follow-up
Studies Composed of Non-Volunteer Students						
(14)* Eckert and Jones	College Freshmen poor high school records	Grade-point average	Control group "not taking course"	Matched on: age, aptitude, curriculum, not on high-school rank	Trained group 1/4 grade-point below control but difference less than high-school Regents exams	Results same for second year; control group persisted longer in college
(19) Kilby	College Freshmen advised to enroll	Average final grade; reading scores	Performance to-tal Freshman class	Matched on: reading scores, grade forecasts, previous achievements	Significant increase grade-point average (from 35th to 50th percentile of class)	--
(19) Kingston and George	College Freshmen	Total grade average	Randomly selected group not taking course	Non-participants had significantly higher ACE scores	--	Two years after 1 of 3 groups showed significant gain
(23)* Mouly	College students with low scores on English test	Grade-point average	Group required to enroll but failing to; students randomly excused	Covariance analysis to allow for differences: reading and ACE scores	Significant increase grade-point average for students finishing course vs. sum of 2 control groups	Non-significant gain at end of 2nd semester
(36)* Willey and Thomson	Freshmen with low IQ's	Grade-point average; reading tests	Group presumably of equal motivation	Matched on: ACE, reading scores	Significant improvement grade-point average (.01 level) and reading scores	--
(34) Tresselt and Richlin	College students with low averages	Study habits score, grades	Own previous average and test score	--	Chi-square significant for 8 of 9 sub-groups on both criteria	--

(37)* Winter	Freshmen lowest 2 deciles intelligence	Grade-point average	One group unable to enroll, second group from previous 2 years	Matched on: intelligence, average high school rank	54% trained group averaged C or above vs. 23% for controls	No difference in 4-yr. grade distributions. Trained group finished .84 semesters more
(13)* DiMichael	9th grade pupils	Achievement tests Latin, algebra, history	Control group given study hall	Matched on: IQ, MA, age, sex, school year, curriculum, teacher	Significant gain only for history; most improvement for middle 2/4ths of ability	--
(15)* Fahey and Waller	High-school students low achievers	Scholastic average	Previous performance; control group randomly excused	Matched on: IQ, disparity between IQ and achievement, age, sex, grade, curriculum	Trained group exceeded control group but not significantly (16 per group); trained group significantly exceeded own previous average	Gains for 3 marking periods
(35)* Turrell	Grades 11 to 14	Grade-point ratios	Control group randomly excused	Matched on: grade, sex, courses, instructors, IQ, previous records and vocational objective	"No evidence all students will improve"; middle-ability groups tend to profit most	Grade-point ratios compared over 4 semesters

semester in which the course is given, so much the better. Information about drop-outs is helpful, but demands care in interpretation since the relation between drop-out-rate and scholastic effectiveness may be obscured by the relation between drop-out-rate and other factors such as finances.

The criterion of scores on reading tests and study-habits inventories may have ancillary value, but needs to be used cautiously for several reasons. First, scores on these tests may not be valid criteria, in the sense that they may not correlate very highly with students' performances at particular institutions. This is borne out by Preston and Tuft's (25) study of Phi Beta Kappa women showing that in some cases these women were very poor readers in terms of usual test norms. Secondly, positive results indicated by these tests may be misleading if regression toward the mean is not taken into account. Deviant sub-groups selected for intensive training, say in reading, on the basis of one administration of a reading test, usually will be less deviant on the second administration of the test, when no conceivable "real" improvement could possibly have occurred (24). If these scores are used even as partial criteria, this regression needs to be discounted in the assessment of results. Also lack of correction for chance of scores on these tests can lead to erroneous conclusions (24).

Finally, it should be remarked that the total grade-point average itself may be misleading because of general improvement of grades as students proceed through college, perhaps owing both to gravitation toward courses for which they have most aptitude and regression toward the mean. Before-after testing of one group is therefore inadequate.

The Baseline - In most evaluations, experimental and control groups were matched on intelligence or scholastic aptitude. It is generally agreed that intelligence and scholastic aptitude should be controlled, or the design should be such that the influence of differing abilities can be determined.

A second important variable, motivation, was not commonly controlled. Only 11 of the 22 studies matched experimental and control groups on desire to take a study-skills course, and 7 of the 11 did this by studying a required course with some excused students providing a control group. For voluntary courses, matching on motivation is difficult, and was accomplished by refusing admittance to some volunteers wishing to take the course, and then collecting control data from these non-admittees (1, 20, 28, 33). Sincerity of the desire to enroll was substantiated by the fact that large numbers of non-admittees later enrolled.

Subjects were often matched on many additional variables, such as age, sex, field of concentration, and initial reading scores where reading formed a large part of the course. Lack of matching on these additional variables (except for reading scores) does not appear to be important in terms of confounding conclusions, but certainly further matching should have the effect of reducing extraneous variation,

and thus permit more sensitive detection of any differences in performance.

Generalization of positive results from the group of students taking a course to an entire student population at a given institution, or to other institutions, may be very risky, because students taking these courses, except the required courses, are probably not very representative of the total student population from which they come. Shaw's (30) investigation of intelligence of students who enrolled in study courses over a two-year period revealed them to be slightly below average. Only one study selected an experimental group that was a stratified sample of the student population, but even here the fact that all students taking the course were volunteers makes generalization to the rest of the student population tenuous.

Results - The main conclusion supported by Table I is that some kind of improvement following a study-skills course seems to be the rule, although the improvement varies from a very slight amount (14) to a considerable amount (20, 33). Among high-school students, those of intermediate ability seem to profit most (13, 35). The consensus of these evaluations agrees with results of Blake's (2) survey: favorable results were noted in all cases where evaluations were undertaken.

When we confine attention to only those studies well-controlled on intelligence and motivation (starred studies in Table I), the range of improvement is just as broad as it is when all 22 evaluations are included. It is noteworthy that it is the required courses which report the smallest gains, although evidence on college-level required courses is scanty and students enrolled in them are atypical. The implication may be that desire to enroll in a course is an indispensable factor for achieving any large improvement.

Smith and Wood's (33) investigation clearly shows the role played by motivation. They included one control group wishing to take the course but unable to schedule it, and a second control group consisted of a representative sample of the Freshman class. The trained group obtained a significantly higher grade-point average than either control group, while the group wishing to take the course but not enrolled scored somewhat higher than the representative sample, although not significantly higher. We can infer from this that motivation per se will not yield a significant improvement unless it is accompanied by participation in a course. The fact that all studies with motivation controlled report significant gains permits us to have some confidence that gains for studies not controlling motivation are probably gains dependent on the courses, rather than gains attributable only to increased motivation. This fortunate state of affairs allows us to include studies not controlling motivation in our general appraisal of the effectiveness of study-skills courses.

Overall judgment about the benefit accruing from these courses needs to be tempered somewhat, in

TABLE II
MAGNITUDE OF GAINS FOLLOWING SKILLS COURSES^{1,2}

Study	Kind of Course	Size of Gain
Barbe (1)	Reading, vocabulary, study habits; 5 hours per week for 12 weeks	About half a letter grade
Charles (11)	Reading, general study methods, library and term paper skills, test-taking skills, diagnostic testing; 10-week laboratory course	0.47 letter grade (?)
Kilby (18)	Reading	15 centile points on class distribution
McDonald (20)	Reading	2.5 percentage points
McGinnis (21)	Reading, general study skills; 16 weeks	0.56 points in point- hour ratio (probably half a letter grade)
Mouly (23)	Reading; 3 hours per week for one semester	0.42 letter grade
Ranson (26)	Reading, individual testing, study habits; varying from part of one semester to three or four semesters	0.43 letter grade
Robinson, H. (28)	"Remedial instruction"; 21 hours over 10 weeks	1.6 percentage points
Simpson (32)	Counseling, reading, "other skills"	0.22 to 0.30 letter grade
Smith and Wood (23)	Reading	0.25 letter grade
Tresselt and Richlin (34)	Lectures on study techniques, group sessions of "free expres- sion," plus individual sessions totaling 2-30 hours	For <u>some</u> sub-groups: 2.3 to 3.3 percentage points
Willey and Thomson (36)	Reading	0.40 letter grade
Winter (37)	Discussion of study methods for 2 hours plus 7 hours supervised study per week	54% averaged C or better vs. 23% for controls
Wittenborn (38)	About two-thirds reading, one- third study methods; 2 hours per week for 7 weeks	About 0.40 letter grade

1. All gains are for a trained group over a control group of equal intelligence
2. College courses only

spite of the uniformly positive results, by awareness that negative results are much less apt to be published than positive results. That there may be negative evaluations unreported is suggested by Christantiello's and Cribbin's (12) survey to which one college responded that two controlled studies had shown a course to be ineffective. Nevertheless, the uniformly favorable results as reported are striking, when one considers the wide variation in kind of course leading to improvement, and the disparate kinds of students enrolled.

Information was collected on the length and content of the courses when it was reported. The length of the courses varied from a total of 7 hours' work (29) to several semesters (27). In some instances effort was devoted exclusively to improving reading skills (18, 19, 21, 30); in other instances, there were lectures and discussions dealing with study habits, individual counseling and diagnostic testing, supervised study of regular course material, practice in the fundamental skills, as well as remedial reading instruction. Eighteen of the 22 studies enrolled college students, (6 of the 18 dealt with persons having poor high-school records or comparatively low intelligence), three enrolled high-school students, and one enrolled medical students.

That improvement follows such diverse courses leads to the conclusion that improvement is likely following many kinds of study-skills courses. The amount of improvement does seem related to whether the course is voluntary or required, although evidence on required courses at the college level concerns students of rather probationary status. All the voluntary college-level courses report gains that are rather impressive, and in every case where follow-up results are available, the gains persist.

The amount of improvement may be statistically significant, but not educationally significant. It is difficult to comment on this point because of the various kinds of grade-scales in use at various schools. Information on the magnitude of the gain was collected for 14 college courses; for the other 4 college courses, no very precise information was available. This information, as well as some indication of the nature of the course, is summarized in Table II. The modal gain is about half a letter grade, and it can be concluded that gains from study-skills courses are often large enough to merit interest.

Conclusions - The evaluative investigations of study-skills courses so far reported seem to bear out the conclusions that:

- (1) A study-skills course will usually be followed by improvement.
- (2) A course will be most beneficial for students desiring to take it.
- (3) Students wishing to take a study-skills course but prevented from doing so, and therefore presumably of comparable motivation to those enrolled, fail to show significant improvement.
- (4) Any gains noted will not necessarily be re-

lated to either the content or the duration of the course.

REFERENCES

1. Barbe, W. B., "The Effectiveness of Work in Remedial Reading at the College Level," Journal of Educational Psychology, 1952, 43, 229-237.
2. Blake, W. S. Jr., "Study-skills Programs," Journal of Higher Education, 1955, 26, 97-99, 114.
3. Brown, W. F., and Holtzman, W. H. "Study-Attitude Questionnaire for Predicting Academic Success." Journal of Educational Psychology, 1955, 46, 75-84.
4. Carter, H. D., "Methods of Learning as Factors in the Prediction of School Success." Journal of Psychology, 1948, 26, 249-258.
5. Carter, H. D., "Correlations Between Intelligence Tests, Study Methods Tests and Marks in a College Course." Journal of Psychology, 1950, 30, 333-340.
6. Carter, H. D., "What are Some of the Basic Problems in the Analysis of Study Techniques?" California Journal of Educational Research, 1951, 2, 170-174.
7. Carter, H. D., "Development of a Diagnostic Scoring Scheme for a Study Methods Test." California Journal of Educational Research, 1955, 6, 26-32.
8. Carter, H. D., "Cross-validation of a Study Methods Test." California Journal of Educational Research, 1955, 6, 32-38.
9. Carter, H. D., "Mechanics of Study Procedure." California Journal of Educational Research, 1958, 9, 8-13.
10. Chahbaz, P., "Analysis of Cornell Orientation Inventory Items of Study Habits and Their Relative Value in Prediction of College Achievement." Journal of Educational Research, 1957, 51, 117-128.
11. Charles D. C., "College Reading and Study Improvement." Journal of Higher Education, 1951, 22, 265-267.
12. Christantiello, P. D., and Cribbin J. J., "Study-skills Problem." Journal of Higher Education, 1956, 27, 35-38.
13. DiMichael, S. C., "The Transfer Effects of a How-to-study Course Upon Different I. Q. Levels and Various Academic Subjects" Journal of Educational Psychology, 1943, 34, 166-175.
14. Eckert, R. E., and Jones, E. S., "Longtime Effects of Training College Students How to Study." School and Society 1935, 42, 685-688.
15. Fahey, G. L., and Waller, C. H., "An experimental Investigation of the Effectiveness of Certain Diagnostic and Guidance Procedures When Applied in Cases of Low School Achievement." Journal of Educational Research

- search, 1941, 34, 335-345.
16. Gough, H. G., "What Determines the Academic Achievement of High School Students." Journal of Educational Research, 1953, 46, 321-331.
 17. Holtzman, W. H., and Brown, W. F., "Study Habits and Attitudes in the Prediction of Academic Success." American Psychologist, 1953, 8, 369.
 18. Kilby, R. W., "The Relation of a Remedial Reading Program to Scholastic Success in College." Journal of Educational Psychology, 1945, 36, 513-534.
 19. Kingston, A. J., and George, C. E., "The Effectiveness of Reading Training at the College Level." Journal of Educational Research, 1955, 48, 467-471.
 20. McDonald, A. S., "Influence of a College Reading Improvement Program on Academic Performance." Journal of Educational Psychology, 1957, 48, 171-181.
 21. McGinnis, D., "Corrective Reading: A Means of Increasing Scholastic Attainment at the College Level." Journal of Educational Psychology, 1951, 42, 166-173.
 22. Michael, W. B., and Reeder, D. E., "The Development and Validation of a Preliminary Form of a Study-Habits Inventory." Educational and Psychological Measurements, 1952, 12, 236-247.
 23. Mouly, G. J., "A Study of the Effects of a Remedial Reading Program on Academic Grades at the College Level." Journal of Educational Psychology, 1952, 43, 459-466.
 24. Murphy, H. D., and Davis, F. B., "A Note on the Measurement of Progress in Remedial Reading." Peabody Journal of Education, 1949, 27, 108-111.
 25. Preston, R. C., and Tuft, E. N., "Reading Habits of Superior College Students." Journal of Experimental Education, 1948, 16, 196-202.
 26. Ranson, M. K., "An Evaluation of Certain Aspects of the Reading and Study Program at the University of Missouri." Journal of Educational Research, 1955, 48, 443-454.
 27. Robinson, F. P., "Two Quarries With a Single Stone." Journal of Higher Education, 1945, 16, 201-206.
 28. Robinson, H. A., "A Note on the Evaluation of College Remedial Reading Courses." Journal of Educational Psychology, 1950, 41, 83-96.
 29. Sharp, S. L., "Effective Study Methods." Journal of Higher Education, 1943, 14, 217-272.
 30. Shaw, J. G., "An Evaluation of a Study-Skills Course." Personnel Guidance Journal, 1955, 33, 465-468.
 31. Sheldon, W., "An Evaluation of an Experimental Reading Program for Medical Students." Journal of Educational Psychology, 1948, 39, 298-303.
 32. Simpson, R. G., "The Reading Laboratory as a Service Unit in College." School and Society, 1942, 55, 621-623.
 33. Smith, D. E. P., and Wood, R. L., "Reading Improvement and College Grades: a Follow-up." Journal of Educational Psychology, 1955, 46, 151-159.
 34. Tresselt, M. E., and Richlin, M., "Differential Prognosis in a College Study Methods Course." Journal of Psychology, 1951, 31, 81-89.
 35. Turrell, A. M., "Study Methods and Scholarship Improvement." Junior College Journal, 1937, 7, 295-301.
 36. Willey, D. S., and Thomson, C. W., "Effective Reading and Grade-point Improvement With College Freshmen." School and Society, 1956, 83, 133-134.
 37. Winter, J. E., "Experimental Study of the Effect on Learning of Supervised and Unsupervised Study Among College Freshmen." Journal of Educational Psychology, 1936, 27, 111-118.
 38. Wittenborn, J. P., "Classes in Remedial Reading and Study Habits." Journal of Educational Research, 1944, 37, 571-586.

AUTHORITARIAN TRENDS IN PERSONALITY AS RELATED TO ATTITUDINAL AND BEHAVIORAL TRAITS OF STUDENT TEACHERS

JOSEPH A. DEL POPOLO

State University of New York, Cortland

THE PURPOSE OF this study was to investigate the relationship between an individual's personality structure and his opinions and attitudes toward pupil-teacher relationships and his observable behavioral traits in a classroom setting. The aspect of personality adopted for the investigation was that of the authoritarian personality structure.

As a test of authoritarianism a 177 item scale was adapted from Webster, Sanford, and Freedman.^{1*} A high score on the scale indicated that a subject was high in authoritarianism. This instrument was referred to as the A Scale.

The measure of attitudes toward pupil-teacher relationships selected for use was the Minnesota Teacher Attitude Inventory, Form A, by Cook, Leeds, and Callis.² This 150 item inventory was designed to measure those attitudes of a teacher which predict how well he will get along with pupils in interpersonal relationships. A high score implied ability to establish harmonious relationships with pupils. This instrument was referred to as the MTAI.

An Observation Check Sheet for recording the classroom behavioral traits of the test subjects was constructed by the author. This 75 item instrument was designed to follow the test subjects into student teaching for the purpose of investigating the relationship of attitudes and actual practice. The higher a score a subject would get on the Observation Check Sheet, the better his relationship with pupils was presumed to be.

The procedure followed employed a pilot study group, an experimental group, and a control group. In all, 366 sophomore and junior students in a New York State teachers college participated in the study. All test subjects were pre-tested on the A Scale and MTAI at the beginning of the sophomore year and post-tested on the same two instruments at the end of the junior year. The Observation Check Sheet was completed for each subject in the pilot study and experimental groups during the student teaching. Subjects in the control group were not assigned to a student teaching experience.

The statistical analysis rendered the following results:

1. Negative correlations of $-.59$ and $-.66$ were found between A Scale and MTAI scores.
2. Negative correlations of $-.62$ were found between A Scale and Observation Check Sheet scores.
3. Positive correlations of $.47$ and $.41$ were found between MTAI and Observation Check Sheet scores.
4. A comparison of the upper and lower fourths on the A Scale with respect to Observation Check Sheet scores produced a t ratio of -6.95 .
5. A comparison of the upper and lower fourths on the MTAI with respect to Observation Check Sheet scores produced a t ratio of 3.35 .
6. A t ratio of 6.52 was found between pre- and post-A Scale scores and 2.01 between pre- and post-MTAI scores for the experimental group.
7. A t ratio of 5.30 was found between pre- and post A-Scale scores and 4.18 between pre- and post-MTAI scores for the control group.
8. Reliability coefficients for the MTAI, A Scale, and Observation Check Sheet were found to be $.97$, $.92$ and $.96$ respectively.

Consideration of the data permitted the drawing of the following conclusions:

The investigation lent support to the main hypothesis that a significant relationship exists between an individual's personality structure and his opinions and attitudes toward pupil-teacher relationships and his observable behavioral traits in a classroom setting.

Authoritarian students tend to get significantly lower scores than equalitarian students on an inventory of attitudes and opinions about pupil-teacher relationships. These differences were interpreted in terms of the dissimilar psychological orientations of the two groups.

Authoritarian students tend to display behavioral traits during student teaching which imply an inability to establish harmonious pupil-teacher rela-

*All footnotes will be found at end of article.

tionships. On the other hand, equalitarian students tend to display behavioral traits which are felt to be conducive toward the establishment of harmonious pupil-teacher relationships.

The student teaching experience with its day-to-day contact with pupils and school problems tends to influence a little the idealistic approach toward children held by student teachers.

Author's Note: A strong relationship is to be expected between observation check sheet and A Scale, since observations of check sheet considered authoritarian acts on the part of the teacher. How-

ever, these acts are revealing of teaching behavior and authoritarian personality jointly.

FOOTNOTES

1. Webster, H., Sanford, N., and Freedman, M., "A New Instrument for Studying Authoritarianism in Personality," The Journal of Psychology, 40: 73-84, 1955.
2. Cook, W. W., Leeds, C. H., and Callis, R., Minnesota Teacher Attitude Inventory. New York: The Psychological Corporation, 1951.

OBSERVATION CHECK SHEET FOR STUDENT TEACHERS

Student's name _____ Teaching Center _____ Grade _____

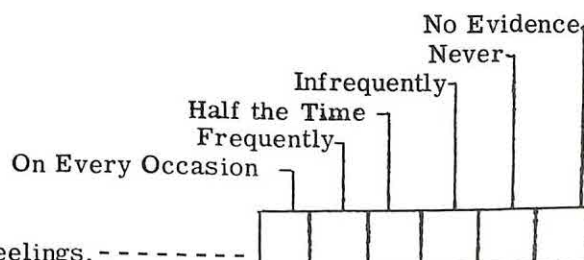
Sponsor Teacher _____ Date _____ 19 _____

Directions: For each item below, place an X in the square which best describes your student teacher's performance, attitude, or behavior during his stay with you.

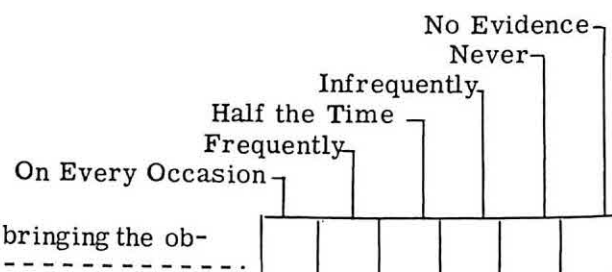
Listed below, you will find brief descriptions of the observation categories to help you indicate your frank and objective observation of your student teacher. The closest approximation is what is wanted.

<u>On Every Occasion:</u>	specific item describes behavior or feeling of student teacher on every occasion the situation occurred.
<u>Frequently:</u>	specific item describes behavior or feeling of student teacher about three-fourths of the time the situation occurred.
<u>Half The Time:</u>	specific item describes behavior or feeling of student teacher about half the time the situation occurred.
<u>Infrequently:</u>	specific item describes behavior or feeling of student teacher about one fourth of the time the situation occurred.
<u>Never:</u>	the student teacher never reacted or behaved in the manner described.
<u>No Evidence:</u>	no situation occurred in which specific item was observable.

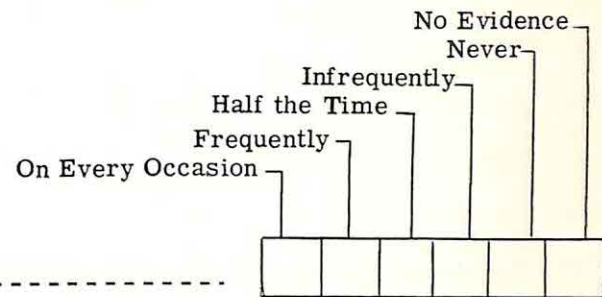
Going from left to right, the squares for the descriptive phrases appear in the same order as that above for items 1-35. For items 36-75, the same phrases are used but the order of the squares is reversed.



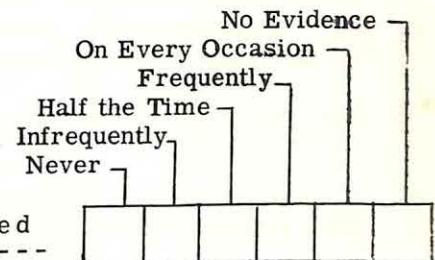
1. The student teacher encouraged pupils to express their own feelings. -----
2. The student teacher gave reasons to pupils for restrictions placed upon them. --
3. The student teacher encouraged and permitted pupils to visit and chat with each other and with him. -----
4. The student teacher gave pupils the opportunity to take over more and more of their decision making. -----



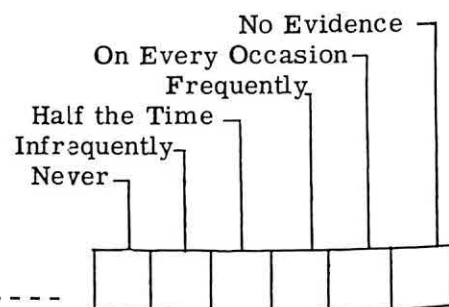
5. When pupils were rebellious, the student teacher encouraged bringing the objections into open discussion. -----
6. The student teacher expected schooling to be enjoyable and school to be a place where children like to come. -----
7. The student teacher turned minor disciplinary problems into jokes. -----
8. The student teacher laughed with pupils when something funny happened. -----
9. The student teacher was more interested in preventive discipline than corrective discipline. -----
10. The student teacher explained requirements and regulations to pupils. -----
11. The student teacher expressed concern about children who obey without question any order or suggestion given. -----
12. In discussion of school practices the student teacher considered the wants of the children a prime concern. -----
13. The student teacher expected each child to have a potential for productive imagination. -----
14. The student teacher allowed for individual differences in pupils in his teaching. -----
15. The student teacher allowed pupils to work independently on class projects and activities -----
16. The student teacher welcomed questions from pupils. -----
17. The student teacher was extremely careful to keep all promises made to pupils. -----
18. The student teacher made efforts to motivate the disinterested pupils. -----
19. The student teacher considered young people today to be just as good as those of the past generation. -----
20. The student teacher re-explained instructions not understood by pupils. -----
21. The student teacher preferred praise to censure as a means of motivation. -----
22. The student teacher provided special material for retarded and advanced pupils. -----
23. The student teacher took an occasional set-back or failure in stride. -----
24. The student teacher allowed pupils to "try things for themselves." -----
25. The student teacher inquired about home conditions of pupils. -----
26. The student teacher adapted his instruction for individual pupils. -----
27. The student teacher gave public recognition to pupils who made the most effort. -



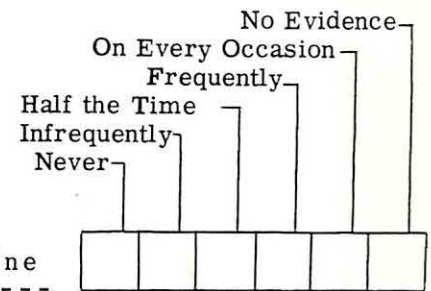
28. The student teacher attended evening functions of the school. -----
29. The student teacher admitted to pupils he didn't know the answer to a question. -
30. The student teacher looked for causes in cases of truancy. -----
31. The student teacher recognized that aggressive pupils are not always the greatest problems. -----
32. The student teacher allowed pupils to select their own topics for study. -----
33. The student teacher helped by doing "extra work" not required. -----
34. The student teacher was cheerful and good-natured with all pupils. -----
35. The student teacher admitted to pupils the mistakes he made. -----



36. The student teacher expressed the point of view that "more old-fashioned whippings" are needed. -----
37. The student teacher required that his opinions must prevail. -----
38. The student teacher was faultfinding about the lack of common courtesies. ---
39. The student teacher demanded rigid conformity on the part of pupils. -----
40. The student teacher considered cheating to be probably one of the most serious of moral offenses. -----
41. The student teacher explained to pupils the need to respect all teachers because of their position as teacher. -----
42. The student teacher interrupted pupils when they were speaking. -----
43. The student teacher discouraged pupils from asking questions about sex. ----
44. The student teacher was disturbed by cultural changes which appeared to give today's children too much freedom. -----
45. The student teacher punished or criticized pupils in front of others. -----
46. The student teacher punished the whole class when unable to locate the cause of the difficulty. -----
47. The student teacher assigned additional schoolwork as a form of punishment. -
48. The student teacher expected pupils at all times to get permission before speaking or leaving their seats. -----



49. The student teacher did not permit whispering in the classroom. -----
50. The student teacher expressed the opinion that teachers generally are too lenient with their pupils. -----
51. The student teacher considered rules and regulations inviolable. -----
52. The student teacher used physical force in disciplining pupils. -----
53. The student teacher criticized publicly pupils with nervous habits. -----
54. The student teacher worried in advance about possible discipline problems. ---
55. The student teacher attempted to be "hardboiled" in his relations with pupils. --
56. The student teacher attempted to supervise and influence companionship of pupils. -----
57. The student teacher for the purpose of motivation made critical comparison of the pupils' work. -----
58. The student teacher gave attention to the impulsive desires and whims of the pupils. -----
59. The student teacher expected most of the pupils to be inconsiderate toward their teacher. -----
60. The student teacher insisted that shy pupils participate in classroom activities. -----
61. The student teacher gave most of his attention to the aggressive pupils. -----
62. The student teacher increased homework assignments for all pupils. -----
63. The student teacher considered pupils generally to be too dependent upon the teacher. -----
64. The student teacher expected pupils to always stand when reciting. -----
65. The student teacher punished or threatened to punish pupils by lowering grades or marks. -----
66. The student teacher was fearful that pupil freedom in the classroom would eventually lead to confusion. -----
67. The student teacher kept pupils after school to prepare lessons they neglected. --
68. The student teacher used marks or grades as motivation for learning. -----
69. The student teacher became bored with teaching. -----
70. The student teacher lost patience with pupils. -----
71. The student teacher criticized modern school procedures. -----



72. The student teacher criticized the parents of pupils who caused discipline problems. -----
73. The student teacher became "upset" or "flustered" while working with slow-learning pupils. -----
74. The student teacher felt that pupil misbehavior is done to annoy the teacher. --
75. The student teacher was unable to control the pupils. -----

READING SKILL AS A PREDICTOR OF COLLEGE ACHIEVEMENT

MALCOLM H. ROBERTSON and MILDRED M. HARRISON
University of Mississippi

CONSIDERABLE emphasis has been placed on the relationship between reading achievement and grades in college. Evidence for the importance of this relationship has been derived from two types of research studies.

The first type is represented by correlational studies in which scores on reading tests are correlated with grades obtained at a later time. Thus, Jackson (4) found a correlation of .64 between the Michigan State College Reading Test and first quarter grade point averages. In another study by Havens (3), a substantial positive correlation was obtained between scores on the Cooperative Reading Comprehension Test and grades in pre-law courses. On the other hand, low or negative correlations have been reported by Preston and Bolton (7) and Murphy and Davis (6).

The second type of research study analyzes the improvement in grades shown after participation in reading programs. For example, Smith and Wood (9) found that students who had taken a reading improvement course showed a significantly greater gain in grade point average than a comparable group of students who wanted to participate but were unable to do so, and another comparable group randomly selected from the freshman population. Similarly, Ranson (8) reported that students in a reading program demonstrated a significantly greater increase in grades than a comparable group of non-participating students. On the other hand, Kingston and George (5) noted a reverse trend in which students who had not taken a reading course made a greater improvement in grades than a group who had completed such a course. However, these authors pointed out that the two groups were not comparable in terms of scholastic aptitude.

The present study follows the design of the correlational type of investigation. The hypothesis is that scores on a standardized reading test show a predictive relationship to college grades.

Procedure

The group studied was composed of 152 freshman students who had participated in a pre-college counseling and testing program prior to registration.

Approximately 56 per cent of the group were male students. The distribution of the students' choices of various schools within the University was as follows: Slightly more than half the students entered Liberal Arts. Of the remaining students, approximately 20 per cent entered the Business School, 20 per cent entered the Education School, and ten per cent the Engineering School.

At the beginning of the testing session, the Diagnostic Reading Test (2), survey section, form H, was administered to the students. The content and difficulty level of the test are intended for grades seven through college freshman year. The survey section has the following timed subtests which yield separate reading scores:

Subtest I: general story-type reading. (15 minutes)

Subtest II: a vocabulary test using words from mathematics, science, English, and social studies. (10 minutes)

Subtest III: paragraph comprehension of subject matter from social studies and science. (15 minutes)

The following scores were obtained from performance on the three subtests:

Score 1 a: words read per minute as recorded by the student on subtest I.

Score 1 b: comprehension of subtest I at the reading rate recorded.

Score 2: vocabulary from subtest II.

Score 3: the sum of scores 1 b and scores on subtest III which is comprehension of work-type material.

Score 4: the total of scores 1 b, 2, and 3, but excluding 1 a (words read per minute).

In the analysis presented in the next section, scores on 1 a (rate), 2 (vocabulary), 3 (comprehension excluding vocabulary), and 4 (comprehension including vocabulary) were correlated with first semester grade point averages. Afterward, the same type of analysis was performed, but the grades in physical education and ROTC courses were excluded from the computation of grade point averages. These courses were excluded because of their non-academic nature and because grades in ROTC courses are based largely on attendance.

TABLE I
CORRELATION BETWEEN GRADES AND DRT SCORES

	Grades			Corrected Grades	
	N	Correction	SE	Correlation	SE
Rate	152	*	--	.26	.08
Vocabulary	152	.43	.08	.46	.08
Comprehension (excl. voc.)	152	.25	.08	.28	.08
Comprehension (incl. voc.)	152	.40	.08	.43	.08

* Inspection of scatter diagram indicated little or no correlation.

TABLE II
EXPECTANCY TABLE FOR GRADES AND VOCABULARY SCORES

Vocabulary Score	Grade Point Average*				Total
	4.00-5.00	3.00-3.99	2.00-2.99	0-1.99	
25-31	10%	10%	24%	56%	12%
32-37	3%	33%	42%	22%	20%
38-46	16%	45%	15%	24%	30%
47-59	46%	32%	15%	7%	38%
	22%	34%	22%	22%	100%

* A = 5, B = 4, etc.

TABLE III
EXPECTANCY TABLE FOR GRADES AND TOTAL COMPREHENSION SCORE

Vocabulary Score	Grade Point Average*				Total
	4.00-5.00	3.00-3.99	2.00-2.99	0-1.99	
33-57	4%	17%	42%	37%	16%
58-67	10%	33%	32%	25%	20%
68-77	15%	45%	18%	22%	30%
78-96	40%	36%	16%	8%	34%
	22%	34%	22%	22%	100%

* A = 5, B = 4, etc.

Results

To test the hypothesis that reading skill is related to grades, product-moment correlations were computed for first semester grade point averages and each of the four DRT subtest scores. The correlations are presented in Table I. The set of correlations on the right hand side of the table are based on corrected grade point averages (excluding physical education and ROTC grades.) Only vocabulary and total comprehension (comprehension score plus vocabulary score) showed substantial correlation with grades.

Using the computed grade point average, the standard error of estimate was corrected for both the vocabulary score and total comprehension scores (comprehension plus vocabulary). The standard error of estimate was .98 for vocabulary and 1.03 for total comprehension. In other words, for about two-thirds of the group the predicted grade point average would be expected to fall within one letter grade of the actual grade point average.

Again, using the corrected grade point average, expectancy tables were set up for both vocabulary scores and total comprehension scores (comprehension plus vocabulary). Table II shows the percentage of students within each grade category for each vocabulary raw score interval. For an individual in the lowest range of scores (25-31), the probability of achieving a grade point average of A or B is approximately .10. On the other hand, the probability of obtaining a grade point average of F is higher than .50. For an individual in the highest range of scores (47-59), the probability of achieving an A or B average is close to .50, whereas the probability of obtaining an F average is only .07.

Table III shows the percentage of students within each grade category for each total comprehension raw score interval. For a student in the lowest range of scores (33-57), the probability of making an A or B average is about .04, whereas the chances of receiving an F average are close to .40. For one in the highest range of scores (78-96), the probability of earning an A average is .40, whereas the chances of having an F average are only .08.

Discussion

There are a number of points to be mentioned in evaluating the results. First, the vocabulary score was the only variable that had a very substantial correlation with grades. Thus, when vocabulary score was added to the comprehension score, the correlation with grades rose from .28 to .43.

Second, the obtained correlation of .46 for the ten-minute vocabulary test is consistent with the correlation of .43 between the 15-minute vocabulary section of the College Qualification Test and the first semester grades. The latter correlation was based on a group of about 600 freshmen at the same institution (1). In addition, a correlation of .57 has been reported for the DRT vocabulary test and the

L and T scores of the ACE (2).

Third, since the standard error of estimate for the two highest correlations was approximately a whole letter grade, the use of these scores for other than a screening purpose is not justified. The expectancy tables indicated that the scores for vocabulary and total comprehension are effective as rough screening devices. Furthermore, the fact that the vocabulary test can be administered in ten minutes adds to the usefulness of such a test as a brief screening measure.

Aside from the above-mentioned points, there are a number of other considerations pertinent to an evaluation of the results. Since the group studied included only one-fourth of the entire freshman group, restriction of range could be an important factor in the size of the correlation. The standard deviation of the group of 152 students, as to grades, was approximately the same as the standard deviation of the total group of 600 freshmen. No data were available on the range of DRT scores for the total freshman group.

Next, the vocabulary test is quite short, and increasing the length of the test might lead to a higher correlation. Even doubling the length of the test would still make it a relatively brief screening device.

The factor of speed probably needs further consideration. Even though the more difficult items have been placed near the end of the test, undoubtedly speed has contributed appreciably to the variance of the scores. Speed might also have contributed to the variance of the grades insofar as the grades might have been based on speeded examinations.

Another factor in the correlations is the reliability of both the predictor and the criterion. While the reported reliabilities for the DRT subtests are at an acceptable level (.80's), the reliability of grading in general is questionable.

Yet another factor is the difficulty level of the test. As mentioned earlier, the difficulty and content level are intended for grades 7 through college freshman year. Perhaps a test with a difficulty level designed primarily for college students would be more appropriate for predicting grades in college.

Finally, as other investigators (5) have indicated, a typical reading test is a better predictor of achievement in courses emphasizing linguistic skill than in other types of courses. In this group, 90 per cent of the students were planning to enter either a business, education, or general liberal arts curriculum.

REFERENCES

1. Bennett, G.; Bennett, Marjorie, Wallace, W., and Wesman, A., *A Manual for the College Qualification Test*. New York: The Psychological Corporation, 1957.
2. Committee on Diagnostic Reading Tests, *Diagnostic Reading Test: Survey Section*. New York:

- The Committee on Diagnostic Reading Tests, 1950.
3. Havens, Virginia, "A Prediction of Law School Achievement from Highschool Rank, Reading Test Scores, Psychological Test Scores, and Average Grade in Pre-law Courses." Journal of Educational Psychology, 1948, 39, 237-242.
 4. Jackson, R., "Prediction of the Academic Success of College Freshmen." Journal of Educational Psychology, 1955, 46, 296-301.
 5. Kingston, A. and George C., "Effectiveness of Reading Training at the College Level." Journal of Educational Research, 1955, 48, 467-471.
 6. Murphy, H. and Davis, F., "College Grades and Ability to Reason in Reading." Peabody Journal of Education, 1949, 27, 34-37.
 7. Preston, R. and Botel, M., "The Relation of Reading Skill and Other Factors to Academic Achievement of 2048 College Students." Journal of Experimental Education, 1952, 20, 363-371.
 8. Ransom, M. Kathleen, "An Evaluation of Certain Aspects of the Reading and Study Program at the University of Missouri." Journal of Educational Research, 1955, 48, 443-454.
 9. Smith, D. and Wood, R., "Reading Improvement and College Grades: a Follow-up." Journal of Educational Psychology, 1955, 46, 151-159.

EXTRAVERSION, NEUROTICISM, AND STUDENT ACHIEVEMENT IN INTRODUCTORY PSYCHOLOGY

A. W. BENDIG
University of Pittsburgh
Pittsburgh, Pennsylvania

The development by Eysenck (8) of the Maudsley Personality Inventory (MPI), which contains 24-item scales of Extraversion and Neuroticism, has stimulated research into the relations between these two personality dimensions and various laboratory measures of learning. However, no evidence is as yet available on the relationship between these scales and traditional measures of academic achievement such as course grades or quality point average, although some indirect evidence suggests that Extraversion is negatively related to achievement while Neuroticism is not correlated with college achievement measures. Eysenck (7) pp. 107-109 has shown that the C, D, and T scales of Guilford's STDCR Inventory (10) are measures of the Neuroticism factor and the R scale measures the Extraversion dimension. Bendig and Sprague (4) found that the Emotional Stability scale (derived from combining the original C and D scales) and the Thoughtfulness scale (derived from the T scale) of the Guilford Zimmerman Temperament Survey (11) were not related to achievement level (grades) in a course in introductory psychology ($N=155$), while the Restraint scale of the GZTS (which is the reverse of the original R scale) was significantly correlated ($r=.20$) with grades. Additional confirmation of the independence of "neuroticism" and course grades can be found in the lack of a significant relationship between Taylor's Manifest Anxiety Scale (12), which is highly correlated ($r=.77$) with the MPI Neuroticism scale (2), and achievement in introductory psychology (1). These results suggest the hypothesis that the MPI Extraversion scale will show a statistically significant negative correlation with course grades, but that the Neuroticism scale will be uncorrelated with this achievement measure.

Neither of the two previous studies referred to above (1, 4) attempted to statistically control for initial differences in verbal aptitude in correlating the temperament scales with course grades. Another study (3) indicated that such a control is necessary in evaluating the predictive usefulness of temperament scales. Edward's Need Achievement scale and Gough's Hr scale showed similar correlations ($r's = .23$ and $.19$) with student quality

point average ($N=164$), but the Hr scale was so highly correlated with a short vocabulary scale also administered to the Ss that the multiple correlation of the vocabulary and Hr scales with QPA was not significantly larger than the zero-order correlation between vocabulary and QPA. However, the combination of both the vocabulary and Need Achievement scales was a significantly better predictor than was the vocabulary scale alone because of the low correlation between the vocabulary and Need Achievement scales. These results suggested that the Hr temperament scale was not a useful predictive instrument when used in conjunction with standard verbal aptitude measures, but that the Need Achievement scale measured an element of college achievement not included in vocabulary scales.

Bendig and Sprague (4) pointed out that another problem in testing the validity of psychometric predictors against achievement criteria is the curvilinear relationship between the average level achieved by Ss on the criterion used and the intra-individual fluctuation of Ss around their level. If course grades are determined by averaging Ss' performance over several course examinations, then an S receiving an A grade must have shown consistent A and B grade performance on course tests. Likewise, an S receiving a failing or F grade must have consistently received F and D grades on course tests. However, some Ss in the middle or C grade range may have consistently received C grades on the course examinations and averaged in the C range while other Ss may have fluctuated widely from A to F on single tests and also achieved in the C range as their average level. It was hypothesized that the mean amount of intra-individual fluctuation on course tests should be smallest for students receiving A and F grades, somewhat larger for Ss receiving B and D grades, and largest for those achieving in the C grade range. This curvilinear relationship was confirmed for two independent samples of students enrolled in introductory psychology (1, 4). In the first study (4) a nonsignificant rectilinear correlation (r) of $-.15$ ($N=155$) was obtained between measures of achievement level and achievement fluctuation while

the curvilinear correlation (η) between these two achievement variables was .31 (significant at the .01 level). The difference between the rectilinear and curvilinear coefficients was also significant at the .01 level. A similar curvilinear relationship ($\eta = .21$) was found for another sample of 352 Ss(1).

The possible importance of intra-individual achievement fluctuation in prediction studies is obvious. The apparent validity of predictors may be reduced because of the relatively low reliability of the achievement criterion due to the presence of inconsistent (high fluctuation) Ss in the achievement level distribution. It was suggested (4) that the inability of temperament measures to add to the predictability of course grades over that achieved by the use of verbal aptitude tests may be due to the presence of this uncontrolled fluctuation variable. If this hypothesis is correct and if either of the MPI scales are related to achievement level, then the validity coefficient of the scale should be higher for low fluctuation Ss (for whom the achievement level criterion measure is more reliable) than it is for high fluctuation Ss.

Procedure

The Ss were 202 undergraduate male students enrolled in six sections of a one-semester introductory psychology course. A plurality of the Ss were sophomores with smaller numbers of freshmen, juniors, and seniors.

Three psychometric scales were administered to all Ss at a single class period within the first two weeks of the semester: (a) a vocabulary test containing 30 multiple-choice synonym items taken from the Cooperative Vocabulary Test, and (b) the Maudsley Personality Inventory (8). Raw scores from these three scales constituted the predictor variables used in this study.

Four 40-item multiple-choice course examinations were administered during the semester and two achievement criteria were derived from these scores. The scores of all Ss on one examination were pooled and converted to standard scores ($M = 75$, $SD = 10$). This procedure was followed for each examination and a letter grade was assigned to each S at the end of the semester by averaging (mean) each S's four standard scores and converting this mean to a letter grade on the basis of previously agreed-upon average standard score ranges: A = 100-85, B = 84-80, C = 79-70, D = 69-60, and F = 59-50. The letter grade received by each S in the course was defined as the criterion of achievement level. These letter grades were quantified in later computations by assigning A grades a value of 3, B grades a value of 2, C grades a value of 1, and D and F grades (because of the small number of F grades) a weight of zero. A measure of achievement fluctuation was obtained by finding the range between the highest and lowest standard scores received by an S over the four course examinations, multiplying this range by the conversion factor given by Dixon and Massey (5, p.

404) to provide an estimate of the variance of the S's examination scores, and extracting the square root of this variance estimate to provide a more normally distributed variable. This is the same measure of achievement fluctuation that has been used in previous studies (1, 4).

The criteria variables of level and fluctuation were intercorrelated by both rectilinear and curvilinear techniques to test the relationship previously found in other samples (1, 4). Each of the letter grade groups was split at the median of the fluctuation measures within that grade group and the four subgroups (one from each of the A, B, C, and combined DF grade groups) with smaller fluctuation measures combined as the "low fluctuation" group of 101 Ss. The four subgroups of Ss with larger fluctuation measures constituted the "high fluctuation" group which also included 101 Ss. The three predictor scales were correlated (product-moment) with achievement level for the total group and also separately for the low and high fluctuation subgroups. The two MPI scales were also correlated with the fluctuation measures within each of the four grade groups.

Results

The rectilinear correlation (product-moment) between achievement level and fluctuation was -.11 (nonsignificant) while the curvilinear correlation (η) between these criterion variables was .26 (significant at the .01 level). The F ratio test of the significance of the difference between these two coefficients showed them to be significantly different at the .01 level of confidence ($F = 8.76$). The trend of the mean fluctuation measures for the four letter grade groups showed the same pattern reported for other samples (1, 4).

Both rectilinear and curvilinear correlation coefficients were computed between each of the three predictor scales and the achievement level criterion for the total group of Ss ($N = 202$). The difference between the r and η coefficients for each predictor was tested for significance by the usual F ratio technique (13), and none of the three F ratios were significant at the .05 level.

The rectilinear (product-moment) correlations between the three predictors and achievement level for the total group and for the high and low fluctuation subgroups can be found in Table I. In addition, Table I also shows the results of t-tests of the significance of the difference between the pairs of correlations for the fluctuation subgroups, the multiple correlation between the Vocabulary and Extraversion of the difference between the Vocabulary-level and Vocabulary-Extraversion-level validity coefficients (13). It can be seen that Vocabulary was positively related to level for the total group and for each of the fluctuation subgroups and that the Vocabulary-level coefficients within the high and low fluctuation subgroups ($r = .44$ and $.43$) were not significantly

TABLE I

ZERO-ORDER AND MULTIPLE CORRELATIONS BETWEEN THREE PSYCHOMETRIC
 SCALES AND STUDENT ACHIEVEMENT IN INTRODUCTORY PSYCHOLOGY

Psychometric Scales	Total Group (N=202)	Fluctuation Subgroups		Significance of Difference (t)
		Low (N=101)	High (N=101)	
Vocabulary (V)	.43**	.43**	.44**	.09
Extraversion (E)	-.17	-.24*	-.11	.98
Neuroticism (N)	-.09	-.01	-.16	1.06
Intercorr. V×E(r)	-.13	-.09	-.16	.51
Mult. Corr. V+E(R)	.45**	.48**	.44**	.31
Signif. of Increase (F)	3.49	5.08*	.18	

*P = .05

**P = .01

different ($t = .09$). Extraversion was significantly related to level for the total group ($r = -.17$) and for the low fluctuation subgroup ($r = -.24$), but not for the high fluctuation subgroup ($r = -.11$). Neuroticism was unrelated to achievement level for all Ss.

The intercorrelations between Vocabulary and Extraversion were not significant at the .05 level within any of the groups of Ss. The Vocabulary-Extraversion combination was not significantly more valid in predicting achievement level than the Vocabulary scale alone for the total group ($r = .43$, $R = .45$) or within the high fluctuation subgroup ($r = .44$, $R = .44$), but the multiple correlation ($R = .48$) was significantly larger (.05 level) than the zero-order correlation ($r = .43$) within the low fluctuation subgroup. Parenthetically it might be noted that the F-ratio test of the difference between r and R within the total group ($F = 3.49$) was significant at the .10 level, but not at the .05 level. The significantly greater validity for the Vocabulary-Extraversion combination with the low fluctuation subgroup was attributable to a slightly lower Vocabulary-Extraversion intercorrelation and a somewhat larger Extraversion-level validity coefficient within this subgroup.

The average correlations within the four achievement level groups between the temperament scales and achievement fluctuation were .00 for the Extraversion scale and -.01 for the Neuroticism scale.

Discussion

The results of this study again confirm the presence of a curvilinear relationship between achievement level and achievement fluctuation. The magnitude of the curvilinear correlation (η^2) between level and fluctuation reported here and in previous studies (1, 4) tends to be moderately stable within different samples of Ss with the reported coefficients being .31 ($N = 155$), .21 ($N = 352$), and .26 ($N = 202$).

The hypothesized relationships between the MPI scales and course achievement were confirmed by the results. Extraversion was significantly and negatively related to achievement for the total group of Ss and the correlation was larger for the low fluctuation subgroup than it was for the high fluctuation subgroup. As previously suggested (4), the validity of the Extraversion predictor was attenuated by the unreliability of the achievement criterion for the total group of Ss. The lack of relationship between the Neuroticism scale and the achievement criterion for the total group, or within either of the two fluctuation subgroups again demonstrates the independence of the neuroticism factor dimension and academic performance. A combination of the Extraversion and Vocabulary scales appeared to be a better predictor than the vocabulary measure alone and suggests that such a temperament scale might be profitably included in future prediction studies. However, the usefulness of Extraver-

sion as a predictor will depend upon the reliability of the achievement criterion and the scale could not be expected to correlate highly with a criterion measure of only moderate reliability.

Summary

The Extraversion and Neuroticism scales of the Maudsley Personality Inventory, were administered, along with a vocabulary scale, to 202 male Ss in introductory psychology. Two achievement measures were obtained for each S: his average performance on four course examinations and his fluctuation on these examinations about his average level of performance. A curvilinear relationship between achievement level and fluctuation was again demonstrated. The Ss were dichotomized within each grade level into high and low fluctuation subgroups and correlations between the three predictor measures and the achievement level criterion were computed for the total group and also for each of the fluctuation subgroups. The vocabulary scale was moderately related ($r = .43$) to achievement level for the total group and also within the fluctuation subgroups. The Neuroticism scale was unrelated to achievement. The Extraversion scale correlated significantly with achievement for the total group ($r = -.17$) and within the low fluctuation subgroup ($r = -.24$), but the correlation within the high fluctuation subgroup ($r = -.11$) was not significant. The multiple correlation of vocabulary and Extraversion with achievement ($R = .48$) was significantly larger than the vocabulary-achievement zero-order correlation within the low fluctuation subgroup.

REFERENCES

1. Bendig, A. W., The validity of two temperament scales in predicting student achievement in introductory psychology. *Journal of Educational Research*, 1957, 50, 571-580.
2. Bendig, A. W., Extraversion, neuroticism, and manifest anxiety. *Journal of Consulting Psychology*, 1957, 21, 398.
3. Bendig, A. W., Comparison of the validity of two temperament scales in predicting college achievement. *Journal of Educational Research*, in press.
4. Bendig, A. W., and Sprague, J. L., The Guilford Zimmerman Temperament survey as a predictor of achievement level and achievement fluctuation in introductory psychology. *Journal of Applied Psychology*, 1954, 38, 409-413.
5. Dixon, W. J., and Massey, F. J., Jr., *Introduction to Statistical Analysis* (2nd. ed.) New York: McGraw-Hill, 1957.
6. Edwards A. L., *Edwards Personal Preference Schedule*. New York: Psychological Corpora-

- tion, 1954.
7. Eysenck, H. J., The structure of human personality. London: Methuen, 1953.
 8. Eysenck, H. J., The questionnaire measurement of neuroticism and extraversion. Rivista di Psicologia, 1956, 50, 113-140.
 9. Gough, H. G., The construction of a personality scale to predict scholastic achievement. Journal of Applied Psychology, 1953, 37, 361-366.
 10. Guilford, J. P., An Inventory of Factors STDCR, Beverly Hills, California: Sheridan Supply Company, undated.
 11. Guilford, J. P., and Zimmerman, W. S., The Guilford-Zimmerman Temperament Survey, Beverly Hills, California: Sheridan Supply Company, 1949.
 12. Taylor, Janet A., A personality scale of manifest anxiety. Journal of Abnormal and Social Psychology, 1953, 48, 285-290.
 13. Wert, J. E., Neidt, C. O., and Ahmann, J. S., Statistical methods in educational and psychological research, New York: Appleton-Century-Crofts, 1954.

THE EFFECT OF SONIC EMOTION ON THE IMMEDIATE RECALL OF NONSENSE SYLLABLES

RICHARD H. BLOOMER and SORRELL E. CHESIN
State University Teachers College, Geneseo, New York

A CONSIDERABLE amount of work has been done relating the affective tone of stimuli to retention. Meltzer (6) and Gilbert (2) have reviewed the literature in this area and find, though the results are not conclusive, that there is a tendency to retain material of positive affect.

It would appear from the fact that humans are interested in sound patterns, such as are found in music and poetry, that the sound of the language is a part of the stimulus complex of a word.

Washburn and Robelee (7) determined that there were differential responses to various language sounds when presented in diphone. Eysenck (5), using Washburn and Robelee's data, determined a number of words which were either positive or negative in sound or both. He then determined the preferences of clinical patients for the words. He found that words which were positive in sound and meaning were liked most. Words which were positive in sound, negative in meaning were liked next most often. Words which were negative in sound and positive in meaning were third preference, and words negative in both sound and meaning were liked least well. Eysenck's study coupled with the evidence in music therapy (3) suggests that positive and negative emotional reactions may be generated as a function of sound alone.

The problem for the present study is first to determine relative positiveness or negativeness for letter sounds, and then to determine the effect of positiveness and negativeness on recall of nonsense syllables.

Procedure

Initial data. Sonic affect scores were determined from judgments of 110 junior college students in California on a two point scale of positiveness or negativeness. The stimuli presented were all of the possible combinations of the forty-two letter sounds of the alphabet which were not meaningful. Two hundred and sixty-seven such diphones were determined. Each diphone consisted of one consonant sound and one vowel sound. Diphones were placed at random in lists of ten each with a pronouncing key consisting of a word which included the diphone

with the significant part underlined. Thirty-one such lists of ten words each were presented to the students. The lists as presented included forty-three repeated diphones as a check for the reliability of the judgments.

All of the students were tested at once. Thirty of them were selected at random and given a list to read to control for voice and tone. The experimenter read the first list of diphones and the students were instructed to listen to the sounds they heard and rate them as positive or negative. Then each of the selected students read his list of diphones. Each of the lists were read in turn and the student who was reading the list was instructed not to respond to the diphones he himself was reading. Thus eighty of the students responded to each of the 310 diphones and the thirty students who read lists responded to all 300 of the diphones save those on their own lists. In this way, 109 responses were received for each of the diphones, and the presentation was controlled so that the experimenter's voice was not a factor in the results.

The data were then placed in a matrix and the sum of the positive scores for each consonant and each vowel sound was found. Since each consonant did not occur with each vowel, each sound was corrected for its occurrence with other sounds in accordance with a method described by Guilford (5, p. 282-292). The corrected scores were then normalized and converted to a 'C' scale as described by Guilford (5, p. 577). The sonic affect score for a given word, then, is the sum of the 'C' scores for each individual letter.

Three procedures were used in determining reliability. First, the above results were correlated with the data presented by Washburn and Robelee (7), which was taken from a similar population forty or more years prior.

Second, the experiment was repeated on twenty-seven freshman students in New York State, and the results correlated with the results from California.

Third, from a combined sonic affect scale, including the results from California and the results from New York, thirty-six diphones were selected at random and made into a tape-recording. These thirty-six sounds were pronounced by six different

individuals, three males and three females. The tape was then played for a second group of twenty-four freshman college students and for two third-grade classes, two fourth-grade classes, three fifth-grade classes and two sixth-grade classes of elementary school pupils, totaling 164 children. These results were then correlated with the results obtained on the tape by the freshman group. Further, these data afforded sex comparisons in relation to responses to letter sounds.

Word sound affect. The validity of the sonic affect scores for words and "Word Sound Affect" or the judgment of musicians as to the relative positiveness or negativeness of the sounds of the words. Musicians for the purpose of the study were twenty-five senior music students at the University of Southern California. These students were asked to rate 3 samples of 498 words, 267 words (on a two point scale) either positive or negative in sound. The results for the twenty-five students were combined and a positive-negative score was achieved for each word. These judgments were termed "Word Sound Affect".

Word Sound Affect score was correlated with the sonic affect scores for the same words. The validity is represented by a partial correlation which holds constant the effects of word length. In addition, correlation coefficients for word-sound-affect and the other variables are presented for comparison with sonic affect.

The Sample. One hundred and four children in eight elementary school classes. The sample for the study was taken from a central school in a town of 7,000.

The stimuli. Two samples of thirty-two nonsense syllables each selected for high, medium, and low sonic affect from Glaze's list (4). List 1 had low association value 0-13%; list 2 had high association value 93-100%, a total of 64 nonsense syllables. The nonsense syllables were then scored for sonic affect and placed at random into 64 lists of eight nonsense syllables each. These served to control the position effects.

Method. Each of the eight classes was presented with eight lists in the following manner. A single experimenter presented all the lists. First the nonsense syllable was read to the class, second the nonsense syllable was then spelled, third it was pronounced again. When the list of eight syllables was presented in this fashion, the pupils were instructed to write as many of the nonsense syllables as they remembered. They were instructed to disregard the order of presentation and just write the ones remembered. The other eight lists were presented in the same fashion.

Data. Only correctly spelled nonsense syllables were counted as correct. The data were corrected for class ability and class size to control further for position effects.

Results of the Study

The sonic affect scale. This is presented in Table I. The letters which are most pleasant are

numbered lower than the letters which are judged least pleasant. Among the most pleasant are the letters 'm' and 'l', both of which are soft sounds, and both of which occur in frequent, highly positive words; the letter 'l' occurring in 'like' and 'love', the letter 'm' occurring in 'mother' and its various derivations, and in 'me'. It will be noted that all of the short vowels except 'ou' and 'u' are rated more positively than average. There is also a tendency for the hard sounds to be rated less positive than average. Notable exceptions to this are 'd' and 'ch'. In no categorization save that of long vowels are the generalizations complete, which would indicate that the affect for a given sound or sound type is a function more of the algebraic sum of its meaning situations in the language, rather than the sum preference for certain kinds of sounds to imply specific meanings.

The sonic affect score for a word represents, then, the sum of the sonic affect score for all the sounds of the word.

Reliability

In each presentation of the lists of diphones in California and in New York, 43 of the diphones were repeated in the lists. Thus there were two scores for each of these diphones; the correlation coefficients between these two presentations for the list read in California is $r = .52$ and for the list read in New York is $r = .58$. Both of these relationships were low but significant beyond the .01 level. It would appear that these correlation coefficients were influenced by the fact that the two presentations of the diphones occurred in differing contexts, were read by different persons, and occurred at different points in the lists.

The preference for letter sounds was determined forty years ago by Washburn and Robelee, and in two places in the United States, nearly three thousand miles disparate. The correlation coefficients between these three measures indicated the consistency of the sonic affect within the culture. The correlation coefficient between Washburn and Robelee's data and the data found in California was $r = .61$. The relation between Washburn and Robelee's findings and the data from New York State was $r = .68$; between California and New York, $r = .54$. While these reliabilities are not as high as those required by tests, they indicate a generalized similarity in sensitivity to sounds on the part of college students.

The relation between the preferences for thirty-six diphones presented on tape to college freshmen and elementary school children of $r = .57$ indicates these two groups respond to the letter sounds in a similar fashion. Elementary school boys and girls respond to the letter sounds in much the same way, as is indicated by the correlation of $r = .81$ between the responses of the two sexes.

It was considered that the sex of the teacher with whom pupils were studying might have an effect on

preference for sound. This was found not to be the case, ($r = .71$) and indicates that responses to the emotional tone of letter sounds are not a function of a particular individual who happens to be currently influencing the child.

These results indicated that sonic emotion reactions to letter sounds are fairly consistent phenomena in our culture, and that therefore some study of their effect on human learning is warranted.

Validity. The validity of sonic affect scores is determined in the present study by the relation between the sonic affect scores for a given word and the judgment of musicians as to the positiveness or negativeness of the sound of the word.

Dividing the data into two groups at random and correlating the sum for each group produced a correlation coefficient of $r = .822$. This indicates that the judgments of the music students were highly reliable. The correlation between sonic affect scores and word sound affect is $r = -.191$. (Table II). The sign was in the expected direction as the word sound affect is a function of a number of positive responses and hence increases with the pleasantness of the sound of the word, whereas the sonic affect score is devised to decrease as the sound of the word becomes more positive. This correlation coefficient is significant beyond the .01 level. The correlation coefficient for the smallest sample was not significant.

These correlations, however, contain the contaminant of word length, which is present in the sonic affect scores. Hence, a partial correlation which holds word length constant gives a truer picture of the relation between sonic affect scores and word sound affect: $r = .333$, $r = .296$, and $r = .329$. Each is significant beyond the .01 level. These correlation coefficients were admittedly low for test criteria, and it is expected that the assumption of serial generalization may account for at least a part of this relationship. However, for present purposes, it was sufficient to say that a relationship exists of sufficient magnitude to warrant further consideration of the variable.

Results. Table III indicated that words of pleasant sonic affect were more likely to be remembered than words of negative or neutral sonic affect. This result was obtained only for words whose association value was high. Nonsense syllables which had a high sonic affect and a high association value tended to be recalled significantly more often than were any other combinations of these characteristics of nonsense syllables. It was interesting to note that significant differences in recall of nonsense syllables occurred only with this particular combination. The correlation between sonic affect scores and the recall of words of low association value was $r = .01$, indicating that no relation existed. On the other hand, the correlation between sonic affect scores and nonsense syllables of high association value were $.42$ which was significant beyond the .01 level. These results indicated that sonic affect was a variable which tended to function in the

recall of more meaningful materials.

Discussion

The evidence presented suggests that sonic affect is a factor in the immediate recall of the more meaningful nonsense syllables. It is interesting to note that the major influence of sonic affect is to be found with nonsense syllables which are highly positive. The difference in recall of nonsense syllables with neutral and negative sonic affect is not significant. Further, there is no significant difference for these two categories of sonic affect between the high association and low association. Thus, for neutral and negative sounding words, the association value is not important. This value becomes significant when the sonic affect is pleasant, and all of the significant differences in immediate recall found in the present study can be attributed to what appears to be a facilitation which accompanies high association value when the emotional tone of the sound of the nonsense syllables is positive. Sonic affect, then, functions as a variable only as nonsense syllables approach meaningfulness. Whether or not this influence remains as important as the reference to the symbols becomes more specific is open to question, but these data suggest that sonic emotion as well as the emotional tone of the meaning of a word may well function in the perception of a word.

It would appear on the basis of this evidence that differential learning of nonsense syllables of low and high association and value is not strictly a function of association alone, but rather a function of association with the emotional tone of the sound. Some of the effects of poetry and music may be attributed to this variable, and perhaps the most practical application of sonic emotion would be in the naming of things and advertising. Further experimentation is suggested to determine the relationship between sonic emotion and the learning of meaningful materials.

Summary and Conclusions

Sonic emotion is defined as the mean judged positiveness of the letter sounds comprising a word. A scale was developed using the judgments of 110 college juniors. The reliability of the judgment was determined by repeating the experiment on college freshmen and third-grade public school children, as well as comparing the data with that of Washburn and Robelee taken forty years previously. Validity was determined by comparing sonic affect scores with judgments of positiveness or negativeness of word sounds by music students. The influence of sonic emotion upon recall was determined by comparing the immediate recall of nonsense syllables of low and high association values and various sonic emotion scores by elementary school pupils. The following results were found.

1. Judgments of sonic emotion are significant-

TABLE I
SONIC AFFECT SCALES

9	8	7	6	5	4	3	2	1
oi	g	f	b	h	a	\bar{a}	\bar{e}	l
	j	n	k	o	e	d	m	
	u	sh	ou	oo	i	\bar{i}	ch	
			t	p	\bar{o}	(y)oo		
			w	r	n	\bar{u}		
			z	v	s			
				y				
				th				
				zh				

* The letter sounds are ranged from that letter most liked (1) in two letter nonsense combinations to that letter least liked in two letter nonsense combinations (9).

TABLE II
CORRELATIONS OF SONIC EMOTION AND WORD SOUND AFFECT

	N	Word Sound Affect
Sonic Emotion	498	-.191**
	267	-.160*
	147	-.084
Partial Correlations with Word Length Constant	498	-.333**
	267	-.296**
	149	-.329**

* r. significant beyond .05 level.
** r. significant beyond .01 level.

TABLE III
MEAN NUMBER OF PUPILS RECALLING NONSENSE SYLLABLES*

Sonic Affect Scores	Number of Syllables Presented	Low Association Value			High Association Value		
		\bar{X}	S^2	t	\bar{X}	S^2	t
1. Pleasant	32	22.27	171.95	1.56(1 and 2)	31.90	359.98	2.74(1 and 2)***
2. Neutral	64	17.80	189.51	.02(2 and 3)	21.12	266.12	.23(2 and 3)
3. Unpleasant	32	17.73	294.61	1.19(1 and 3)	20.20	375.24	2.44(1 and 3)**

*N = 104 **t_{.05} = 1.96 ***t_{.01} = 2.58

ly correlated through time, space, and age of judge.

2. The relationship between sonic emotion scores and judged emotional tone of words is significant.

3. Nonsense syllables of high sonic affect and high association value are recalled significantly more often than nonsense syllables of neutral or low sonic affect, or nonsense syllables low in association value, regardless of the sonic emotion. From these data we can infer that differences in recall of nonsense syllables of low and high association value are a function of sonic affect in combination with high association value.

REFERENCES

1. Eysenck, N. J. "Abnormal Preference Judgments as Complex Indicators," *American Journal of Orthopsychiatry*, 12, 388-45, 1942.

2. Gilbert, G. M., "The New Status of Experimental Studies on the Relationship of Feeling to Memory," *Psychological Bulletin*, 35, 26-35, 1938.
3. Gilman and Paperte, Music Research Foundation, *Music and Your Emotions*, Liveright, New York, 1952.
4. Glaze, J. A. "The Association Value of Nonsense Syllables," *Journal of Genetic Psychology*, 35, 255-67, 1928.
5. Guilford, J. P., *Psychometric Methods*, Second Edition, McGraw Hill, New York, 1954, p. 597.
6. Meltzer, H., "The Present Status of Experimental Studies on the Relation of Feeling and Memory," *Psychological Review*, 37, 124-39, 1930.
7. Washburn, N. F. and Robelee, L., "The Affective Value of Articulate Sounds," *American Journal of Psychology*, 23, 579-83, 1912.

PRINCIPLES OF CHILD GROWTH AND DEVELOPMENT*

CHARLENE D. PALMER
University of California Santa Barbara
Goleta, California

Introduction

The accumulation of a large body of scientific information in the field of child growth and development presents a serious problem to the teacher who must select the content for a course in homemaking. Although this is a comparatively new field, research comes from many disciplines, such as medicine, psychology, home economics and education. Enrollment of one and a quarter million students in various homemaking courses in grades nine through twelve, during 1950(1), signifies the importance of the problem. The extensive knowledge collected, the large number of students reached, and the limited time devoted to child growth and development in homemaking courses indicate a need for the curricular content to be based upon something more scientific than the opinions of those who develop the courses of study.

Statement of the Problem

This study sought to determine some objective bases for the selection of the content in the area of child growth and development by identifying the basic principles. Outstanding people in the field of curriculum construction agree that one way to determine the curriculum in a field is by locating fundamental principles. The use of principles as a basis for the content of a course in any field is supported by leaders of several disciplines. Their belief is that the facts may change but the principles will change little, if any.

Method of Investigation

A jury of eight people nationally known in the field of child growth and development agreed that seven of eleven books, selected by the writer, were authoritative sources of principles. A content analysis was made of these "approved" books. If a principle was agreed upon by four or more of the seven

books analyzed, it was included. In this manner sixteen principles were agreed upon.

Principles of Child Growth and Development

Found in Seven "Approved" Books

Behavior patterns of children change with maturity.

Development occurs at different rates for different parts of the body. The arms, for example, quadruple their length; whereas the legs quintuple their birth length.

Development progresses in a cephalocaudal direction, that is, from the head downward. For instance, in the infant skin sensitivity comes in the uppermost part of the body before it appears in the lower.

Development also proceeds in the proximodistal direction, which means that motor development occurs earlier in the structures lying nearest to the main axis than those in a more remote area. Muscle control, for example, appears sooner in the arms than in the fingers.

Development proceeds from general to specific responses. For example, at first an infant shows his happiness by total bodily expression. When he grows older, he responds with a smile.

As the child grows, his expressional behavior becomes more refined. At eighteen months the child may throw a temper tantrum when another child takes his toy. When he is four years old, he will probably call the other child names.

Each developmental stage has certain traits characteristic of it. (The traits upon which there was agreement among the books analyzed are listed in the following section.)

Every child normally passes through each of several major stages of development.

Most traits are correlated in development. The child whose intellectual development is above average is generally above average in health, size, so-

* This article is based upon portions of the author's unpublished doctoral dissertation entitled A Secondary School Curricular Investigation on Child Growth and Development, Dr. Ralph K. Watkins, Advisor, University of Missouri, 1957.

ciability, and special aptitudes.

The child develops as a unified whole. His intellect is related to his physical well-being; his physical health is affected by his emotions; his emotions are influenced by school success or failure, by his physical health and by his intellectual capacity.

Both the rate and pattern of the growth of the child can be modified by forces within and without the body.

Child growth is both quantitative and qualitative. The baby's digestive tract is an illustration of this principle. Growth in size permits a larger intake at a single feeding. Changes in structure permit the digestion of more complex foods and increase the efficiency in converting foods into simpler forms which the body can use.

Child growth follows an orderly sequence. For the great mass of children, learning patterns, for example, follow each other in a fixed order. A child learns to stand before he walks, and he passes through the babble stage of syllables in language before he speaks clearly.

Child growth is a continuous process.

Child growth is not even in tempo. There are periods of accelerated growth and periods of decelerated growth. During infancy and the early preschool years growth moves rapidly, and the maturity indicators of each of the various aspects of growth appear in swift succession. During the later preschool and school years the rate of growth decreases although significant changes are taking place. Before puberty certain phases of growth are accelerated before they taper off to the adult level.

Growth patterns are not uniform from child to child.

"Ages and Stages"

On two principles, "Each developmental stage has traits characteristic of it," and "Every child normally passes through each of these major stages of development," the authors of the "approved" books had almost complete agreement. As to exactly what specific "ages and stages" these principles applied was another problem. The authors seemed to agree on the ages when the child is likely to walk, talk, and develop similar abilities; however, there seemed to be agreement upon only the following "ages and stages" which are generally regarded by the layman as common problems.

Somewhere between six and twelve months the infant has a characteristic behavior problem of putting fuzz, pins, hair, dirt, and the like in his mouth.

Negativism, a form of behavior in which the child shows resistance to adult authority, is common to children from eighteen months to three years. (Research substantiating this "age and stage" was conducted with children living in the United States. Some sociologists have found that children living in certain other cultures do not display negativism at this age.)

Concurrently with the negative stage is the exploratory stage. One stage greatly affects the other here. The child is intent upon exploring, and as a result is reprimanded constantly in some environments. So the number of reprimands may affect the degree of negativism.

Handling the genital organs is part of the general exploratory behavior characteristics of the child from six months to five years.

Group participation in play activities does not begin at once. A two-year-old participates in parallel play. By the time the child is three years old, he learns to take turns and cooperate. Some three-year-olds may begin parallel play with other children if they have had no experience in playing with children. Every child goes through these stages of development although the ages may vary depending upon the social development of the child.

The period of physiologic anorexia, the loss of appetite due to a slower growth rate, often appears during the second and third years and should be anticipated and recognized as such.

Stuttering is quite common at around two and one-half years. There was no agreement as to the cause although many causes were given. It was agreed that different children stutter as the result of different causes.

Preschool children have imaginary companions when they are around three and four years old. There was no common agreement as to why they create these companions.

Three and four-year-olds are at the "question-asking" stage. No explanation of this was given in any of the books. The authors may have assumed that the reasons for this stage of development were common knowledge.

Recommendations

The incorporation of the principles of child growth and development into the homemaking curriculum is based upon a readiness organization. (7) At the same time, the content is in accordance with several beliefs about the purposes of education. From these beliefs the bases for the selection of experiences for a curriculum have evolved. The preparation for future activities, the improvement of present activities, the development of obligations and responsibilities, and the accomplishment of the developmental tasks are some of the bases which have developed in the in the order mentioned, according to Williamson and Lyle. (8)

Since research indicates that learning and retention of facts are greater when knowledge is organized and related rather than specific and isolated, the principles may be used as a guide in organizing and relating the content of a course in child growth and development. Facts can be learned in relation to these principles and, as a result, can be retained best in this relationship. When learning occurs in this manner, it is possible for an individual who remembers the principle to proceed relatively easily to other

facts included under the same principle. (2)

Since the working capital of any individual consists of his existing experiences, the application of the principles to be taught should be within the range of experiences of the learner. For instance, most adolescents have heard parents' excitement about the first step taken by their offspring. Also, some students have had a prized possession destroyed by a younger brother or sister. The following principles apply to these two situations and may help the students understand the behavior of the child.

Behavior patterns of children change with maturity.

Each developmental stage has certain traits characteristic of it.

Every child normally passes through each of several major stages of development.

The child develops as a unified whole.

Child growth follows an orderly sequence.

Additional occurrences described by class members give them an opportunity to find some possible reasons for the behavior exhibited by the child. In this way, the principles can be applied to other situations. Additional principles may be introduced too. This application-to-principles organization, a type of readiness organization, is called psychological organization by educational psychologists.

Since many concrete phenomena subsume under a principle, the inclusion of all of the facts in one course is an impossibility; therefore, some specifics need to be selected. The organization of a course around problems, another readiness organization, facilitates the selection of these facts. Since personal problems seem to have reality and significance for adolescents (6), the principles of child growth and development may be taught by studying sibling relationships and baby sitting. Understanding and getting along with brothers and sisters are recognized by both boys and girls as very important problems of high frequency. (3) Baby sitting, for children from infancy to middle childhood, is one of the teenager's most popular means of earning money; and caring for children presents many problems. By using the principles of child growth and development to solve these current problems the student may learn to apply these principles.

Since certain basic laws run through all divisions of data collected in child growth and development (5), students may use these same principles to solve the problems which will probably confront them in the future as mothers and fathers.

It is hoped that the teen-ager will understand to some degree the importance of his contribution to his family - that his behavior has an effect upon the relationship of the family members. Perhaps an awareness of the parents' influence upon their chil-

dren, and, consequently upon society as a whole, will help these adolescents feel a greater responsibility when they are privileged to become mothers and fathers.

During the period of adolescence each individual is expected to attain certain developmental tasks (4); for the adolescents in the United States one of these developmental tasks is preparing for marriage and family life. Although family factors basically determine this attainment, educational institutions can supplement (not supplant) the home by planning the total school curriculum to incorporate subject matter which will aid the accomplishment of these goals. In order to do this, child growth and development should be an integral part of the adolescent's school program, thereby contributing toward the completion of one of the developmental tasks.

Although there are several common types of readiness organizations, only two, applications to principles and problem, are recommended herein. The selection of experiences to be included in child growth and development may be considered in agreement with several beliefs as to the purpose of education.

REFERENCES

1. "Biennial Survey of Education in the United States," Offerings and Enrollments in High School Subjects, 1948-50, Federal Security Agency, Chapter V (Washington, D. C.: United States Printing Office, 1951), pp. 22-23.
2. Bloom, Benjamin S. (ed.). Taxonomy of Educational Objectives (New York: Longmans, Green and Company, 1956), p. 35.
3. Dales, Ruth J. A Method of Measuring Developmental Tasks at the Beginning of Adolescence (Unpublished doctoral dissertation, Cornell University, 1953), 165 pp.
4. Havighurst, Robert J. Developmental Tasks and Education (New York: Longmans, Green and Company, 1952), pp. 52-54.
5. Olson, Willard C. Child Development (Boston: D. C. Heath and Company, 1949), p. 67.
6. Romine, Stephen A. Building the High School Curriculum (New York: The Ronald Press Company, 1954), p. 121.
7. Watkins, Ralph K. Techniques of Secondary School Teaching (New York: The Ronald Press Company, 1958), pp. 57-70.
8. Williamson, Maude, and Mary Stewart Lyle. Homemaking Education in the High School (New York: Appleton-Century-Crofts, Incorporated, 1954), pp. 68-72.

PLACEMENT BUREAU DATA AND TEACHING SUCCESS

MAURICE F. FOSS
Miami University
Oxford, Ohio

College placement bureaus frequently prepare a folder containing a variety of information about prospective teaching candidates. In a few instances employing officials have voiced the opinion that some of this material is misleading.

During the course of an intensive investigation of a selected group of thirty-three industrial arts teachers certain points regarding the relation of placement bureau data and actual teaching performance were discovered. During this study, these teachers were observed in the classroom. The administrators were interviewed, and a thorough study of placement bureau and other campus data was made.

On the basis of their teaching performance, the teachers were divided into three groups judged as superior, satisfactory, and least-satisfactory. The chart presents contrasting items of placement bureau data for the most and less-successful teachers.

Sources of Placement Bureau Data.

There were five sources of placement bureau data examined:

1. Instructor ratings. These ratings are compiled during the candidate's senior year. The checklist contains twenty-two items to be rated A, B, C, D, or unknown. These are rated by instructors who have had the student in one or more courses during his four years of college.
2. Student teaching report. The form of this report has varied somewhat over the period of time studied. The current form calls for a written evaluation of the student's activity, and apparent success. It suggests specific items which should be included in the report.
3. Philosophy of teaching. The student's philosophy of teaching is recorded during his senior year. This is a free expression of his views by the student.
4. General recommendations. These are usually written by three members of the prospective teacher's home community, although in some cases, by a former employer. In all instances, these recommendations were unstructured.
5. Course and grade record. A photostat of the four-year record is included in the student's folder. This provides a list of all courses taken and

grades earned in each.

Comparison of Data for Superior and Less-Satisfactory Teachers.

The extent to which specific placement bureau items distinguish between the superior and less-satisfactory teachers is presented in tabular form. On the left are typical words and phrases used in the ratings or references, and on the right are indicated the relative frequency of use in describing the two groups. The sources of the data are indicated by the center headings. (See next two pages.)

Supplemental Comments on Comparisons.

In the entire group of teachers studied there were only two who had not done any student teaching in industrial arts. Both of these teachers were having considerable difficulty in their teaching, although only one was a beginning teacher. Two other teachers who had done student teaching in driver education, but not in a shop situation, were experiencing a great deal of difficulty -- again, a first- and a third-year teacher. Administrators owe it to themselves and to their students to employ only adequately prepared teachers.

It is worth noting that all superior teachers had general recommendations in their placement bureau folder; but three of the less satisfactory teachers lacked this information. The sources of general recommendations were interesting. Half of those for the superior teachers came from school personnel or ministers, while only about one-fifth of the recommendations for the less-satisfactory teachers came from these groups. While the phrases used to describe the candidates undoubtedly reflected their standards and ideals, the points mentioned did have a relationship to the success of the teachers.

There was not always agreement among those instructors who furnished senior references. If a student had an unusual amount of work in a given area, and had done well in it, that instructor was apt to rate him higher on general personality traits than were his other instructors.

Items in Placement Data Sources,
and Relative Frequency of Use

Words or Phrases	Comparative Ratings of Superior and Less-Satisfactory Teachers.
<hr/> Instructor Ratings <hr/>	
Senior rating on twenty-two items such as "Intellectual ability," "Fluency," and "Probability of growth."	Most ratings were A or B with a few C ratings, and none below. While superior teachers averaged slightly higher than the less-satisfactory teachers, there was little difference between the groups.
<hr/> Comments Added by Instructors <hr/>	
Words or phrases describing personal relations with other students, i. e., "gets along well with others," "agreeable manner."	Applied equally to both groups.
Words and phrases describing mental characteristics; i. e., "able," "clear thinker," "high technical skill."	Used nearly as often to describe less-satisfactory groups as superior groups.
Words and phrases describing attitudes toward work i. e., "dependable," "steady," "takes work seriously," "assumes responsibility."	Used three times as often with superior groups as with less-satisfactory group.
<hr/> Student Teaching Reports <hr/>	
Phrases describing relation of teachers to class; i. e., "creates enthusiasm and interest."	Used five times as often with superior teachers as with less-satisfactory teachers. Reported that students liked superior teachers; no mention regarding less-satisfactory teachers.
<hr/> Philosophy of Teaching <hr/>	
Phrases indicating an interest in children.	Mentioned four times as often by superior teachers.
Phrases indicating an interest in society and a belief in the school as an instrument of society.	Mentioned nearly four times as often by the superior teachers as by the less-satisfactory teachers.
Phrases indicating an interest in subject matter.	Mentioned nearly as often by the less-satisfactory teachers as by the superior teachers.
Phrases mentioning monetary rewards and prestige of teaching.	Mentioned twice as often by the less-satisfactory teachers as by the superior teachers.
<hr/> General Recommendations <hr/>	
"likes children," "interest in children."	Mentioned with superior teachers but not with less-satisfactory teachers.
"honest," "conscientious" "energetic," "enthusiasm"	Used four times as frequently with superior groups as with less-successful groups.

Words or Phrases

Comparative Ratings of Superior and Less-Satisfactory Teachers

Persons referring to the candidate's success in high school; i. e., "upper third of class," "upper fourth of class," "good grades," "intelligent."

Used four times as often with the superior group as with the less-satisfactory group.

Phrases indicating ability to get along with others: "cooperative," "friendly."

Mentioned more than two times as often for the superior group.

Desirable work habits and attitudes such as: "seriousness of purpose," "willing to work."

Mentioned twice as often in connection with superior teachers as less-satisfactory teachers.

Items describing general bearing: "has poise" "can adapt to situations."

Mentioned twice as often with superior teachers as with less-satisfactory teachers.

General characteristics such as "sense of humor" "good moral character," "neat appearance."

Mentioned about equally for each group.

Grade and Course Record

Undergraduate majors and minors.

All superior teachers had majors or double majors in industrial arts.

Master's degree

Three of the less-satisfactory teachers had minors only in industrial arts.

Grade average

With nearly equal years of teaching experience, four times as many superior teachers had earned the master's degree as had the less-satisfactory teachers.

Superior teachers made better grade averages than did less-satisfactory teachers, although the margin narrowed as they advanced through college.

	Superior	Less-Satisfactory
First semester	2.710	2.245
First year	2.762	2.369
Fourth year	2.875	2.612
Industrial Arts	3.383	3.165
Professional	2.620	2.568

Grade requirements for student teaching and graduation would tend to force grades upward. This means that while the superior teachers started initially with a better grade record than did the less-

satisfactory teachers, merely by remaining in college and becoming eligible for student teaching and graduation, the averages by the poorer teachers had to rise.

ARE PRESENT READING TESTS VALID FOR BOTH GIRLS AND BOYS? *

WALTER J. PAUK **
Cornell University
Ithaca, New York

Purpose of Study

The purpose of this study is to find out what some of the specific skills and attitudes, as measured by tests other than reading tests, are possessed by groups of boys and girls who have been categorized into above-average and below-average reading groups. Perhaps data on these types of skills might furnish some additional clues in helping to devise a reading test to clarify the problem on sex differences in reading skills (2, 3, 4).

Sample and Procedures

A battery of tests was administered to 300 ninth-grade students (150 boys and 150 girls). The Cooperative Reading Comprehension Test, C1 (Lower Level) was used as the criterion for determining reading achievement.

To enhance the probability of establishing a clear differentiation between above- and below-average readers, approximately 47 per cent of the boys and girls whose scores fell in the middle range were eliminated by using the probable error of .675 of a standard deviation for the Cooperative Reading Comprehension Test scores.

The students were thus divided into four categories consisting of (a) 42 above-average boy readers, (b) 38 below-average boy readers, (c) 43 above-average girl readers, and (d) 37 below-average girl readers.

Preliminary Statistical Analysis

The original test battery consisted of 28 tests and subtests. Since it was not feasible to analyze all of these variables, the tests which yielded non-significant point biserial coefficients of correlation were eliminated. To obtain additional evidence of relationship between pairs of test scores, a product-moment coefficient of correlation was computed,

and the tests (listed below) which appeared most representative were chosen for further statistical analysis.

- a. Reflective Test of Thurstone Temperament Schedule.
- b. Brown-Holtzman Survey of Study Habits and Attitudes.
- c. Numerical Ability Test of the Differential Aptitude Tests.
- d. Abstract Reasoning Test of the Differential Aptitude Tests.
- e. Spelling Test of the Differential Aptitude Tests.
- f. Sentences Test of the Differential Aptitude Tests.
- g. L-Score of the A. C. E. Psychological Examinations.

Final Statistical Analysis

Discriminant functions (1) with seven line variables for each group were solved, and the data (see Table I) revealed the following relative contributions for each variable for girls and boys.

Conclusions

An inspection of Table I reveals that the language-type tests; namely, the L-Score (ACE), Sentences (DAT), and Spelling (DAT) contributed 80.90 per cent to the total test battery in differentiating the above- and below-average girl readers; whereas, these same language-type tests contributed only 37.49 per cent to the total test battery in differentiating the above- and below-average boy readers.

Most startling were the contributions made by the Abstract Reasoning Test and the Numerical Ability Test. In differentiating the above- and below-average girl readers, these two tests contributed only 11.57 per cent; whereas, these two tests contributed 43.28 per cent to the total test battery in

* Based on doctoral dissertation completed in September, 1955, at Cornell University.

** Director, Cornell Reading Improvement Program

TABLE I

CLASSIFICATION OF RELATIVE CONTRIBUTIONS OF SEVEN
LIKE VARIABLES FOR GIRLS AND BOYS

Test	Relative Contribution	
	Girls	Boys
L-Score (ACE)	46.05	29.71
Spelling (DAT)	17.23	7.11
Sentences (DAT)	17.62	.67
Abstract Reasoning (DAT)	6.45	19.59
Numerical Ability (DAT)	5.12	33.69
Study Habits (Brown-Holtzman)	3.77	6.90
Reflective (Thurstone)	3.76	2.33
	100.00	100.00

differentiating the above- and below-average boy readers.

These results seem to show that certain specific skill and attitude tests differentiate above- and below-average girl readers, and above- and below-average boy readers when the Cooperative Reading Comprehension Test, C1, lower level is used as a measure of reading ability. The results are such that further research along these lines might be fruitful.

...o...

REFERENCES

1. J. S. Ahmann, "An Application of Fisher's Discriminant Function in the Classification of Students," *The Journal of Educational Psychology*, 3: 184-188, 1955.
2. F. B. Davis, "What Do Reading Tests Really Measure," *English Journal*, 33: 180-187, 1944.
3. W. S. Gray, "Needed Research in Reading," *Elementary English*, 29: 100-8, 1952.
4. A. G. Wesman, "Separation of Sex Groups in Test Reporting," *The Journal of Educational Psychology*, 4: 223-229, 1949.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 53

APRIL 1960

NUMBER 8

CONTENTS

✓ Factors Affecting the Development of Cooperation in Children —Alex F. Perrodin.....	283
Measuring Attitudinal Outcomes of Child Psychology with the Parental Attitude Research Instrument—Frank Costin	289
✓ A Comparison of the Problems of Certain Anglo- and Latin- American Junior High School Students—Paul Witherspoon.....	295
Efficacy of Two Tests in Differentiating Potentially Low from Average and High First Grade Achievers—Elmer F. Morgan.....	300
Publication Activities of Members of the Division of School Psychology of the American Psychological Association— Reginald Lanier Jones.....	305
Children's Preferences in Picture Story Book Variables— Ruth Helen Amsden	309
A Survey of Educational Disability in Emotionally Disturbed Children—Arthur S. Tamkin.....	313
A Study of Practices and Provisions for the Gifted Pupil in Mathematics—Monte Scott Norton	316
Book Reviews.....	318

•••••

DEMBAR PUBLICATIONS, INC.

Madison 3, Wisconsin

FACTORS AFFECTING THE DEVELOPMENT OF COOPERATION IN CHILDREN

ALEX F. PERRODIN
University of Georgia

TEACHERS ARE constantly making judgments as to the degree of cooperation manifested by their pupils. When children are cooperative, teaching is a most enjoyable task. But what factors affect the development of this trait so highly characteristic of a democratic citizen? To what extent do various factors of home, community, and school life affect the development of cooperation in a child?

In an attempt to determine the possible relationships of certain factors of home, community, and school life to the development of cooperation in children, a study was made of the 352 children enrolled in fourth through seventh grades of an elementary school. The children in this school located in a small southern city came from mixed socioeconomic backgrounds.

Procedure

Data reported in this study were derived from a comparison of ranks on cooperation earned on the Behavior Preference Record (by Hugh B. Wood, published by the California Test Bureau, 5916 Hollywood Blvd., Los Angeles 28, California) and responses to a personal data sheet consisting of 37 items of home, community, and school information.

The Behavior Preference Record is an evaluation instrument on which boys and girls in grades 4-12 may indicate their knowledge of, and preference for, certain kinds of social behavior in typical home, community, and school situations. In addition to cooperation, this test attempts to evaluate the traits of friendliness, integrity, leadership, and responsibility. Although all of these characteristics were studied with relation to items on the pupil information sheet, this article is limited to the study of factors affecting children's preferences for cooperative behavior.

Cooperation is defined by the Behavior Preference Record as a characteristic of democratic behavior in which the child expresses a preference to deal with others in an adaptive, conformative, and helpful way.

The personal data sheet used collected information as to the child's grade level, sex, age, father's and mother's employment, sibling relationships, length of residence in community, in present house,

in present school, attendance at church, scout groups, "Y" groups, ballroom dancing instruction groups, popularity in peer group, leadership experience as class officer, representative on school council, or service on school patrol, TV viewing habits, reading habits, and reading achievement.

An IBM tabulating machine was used to facilitate the handling of the data obtained from these two instruments.

Findings

Table I indicates that sixth graders tended to score lowest and seventh graders highest in cooperation preferences. A further study of the 352 children by age groups revealed no pattern of growth or decline in cooperation from age nine through age fourteen, although the thirteen-year-olds scored significantly higher than other age levels.

Boys tended to score higher than girls in preferences for cooperation, as indicated in Table II. This finding is contrary to the usual teacher judgment of actual behavior and that reported by Perkins (4, p. 228).

An examination of certain family and home factors revealed no significant differences in ranks in cooperation in relation to the fathers' employment. Those children whose fathers were self-employed or employed by the university tended to be slightly higher in cooperation than those children whose fathers were not self-employed.

Approximately forty percent of the children's mothers were employed full or part-time outside the home. This factor and the factor of thirty-one percent of the mothers not being home when the child returned from school in the afternoon had no significant relationship to the child's rank in cooperation.

Table III indicates that the youngest or younger child in the family tended to score higher in cooperation than other siblings or the child without brothers or sisters.

Children who rated very high or high in the cooperation portion of the Behavior Preference Record tended to be those who had lived in the same house

TABLE I
DISTRIBUTION OF RANKS IN COOPERATION BY GRADES

Rank	Grades									
	Four		Five		Six		Seven		Totals	
	N	%	N	%	N	%	N	%	N	%
Very High	3	3.1	1	0.9	4	5.3	12	16.0	20	5.7
High	29	30.2	40	37.7	16	21.3	23	30.7	108	30.7
Average	43	44.8	45	42.5	28	37.3	34	45.3	150	42.6
Low	16	16.7	11	10.4	12	16.0	4	5.3	43	12.2
Very Low	5	5.2	9	8.5	15	20.0	2	2.7	31	8.8
Totals	96	100.0	106	100.0	75	99.9	75	100.0	352	100.0

TABLE II
DISTRIBUTION OF RANKS IN COOPERATION BY SEX

Rank	Boys		Girls		Total	
	N	%	N	%	N	%
Very High	13	6.7	7	4.4	20	5.7
High	69	35.8	39	24.5	108	30.7
Average	72	37.3	78	49.1	150	42.6
Low	26	13.5	17	10.7	43	12.2
Very Low	13	6.7	18	11.3	31	8.8
Totals	193	100.0	159	100.0	352	100.0

TABLE III

DISTRIBUTION OF STUDENTS RATING VERY HIGH OR HIGH
IN COOPERATION IN RELATION TO HOME FACTORS

Factor	Rank in Cooperation			
	Very High		High	
	N	%	N	%
1. Father is:				
a. self-employed	3	7.9	12	31.6
b. employed by another	12	4.8	72	28.9
c. employed by university	1	3.5	11	37.9
d. in armed services	2	11.1	5	27.8
2. Mother is:				
a. employed away from home	7	5.0	45	32.1
b. not employed away from home	13	6.1	63	29.7
c. usually home in p. m.	15	6.1	75	30.7
d. usually away in p. m.	5	4.6	33	30.6
3. Child is:				
a. oldest or older child	11	7.9	31	22.3
b. middle child	3	4.8	21	33.3
c. youngest or younger child	5	4.9	42	40.8
d. an only child	1	2.1	14	29.8
4. Child has lived in community:				
a. all of life	9	4.6	63	32.1
b. more than half of life	4	6.0	22	32.8
c. less than half of life	7	7.9	23	25.8

TABLE IV

DISTRIBUTION OF STUDENTS RATING VERY HIGH OR HIGH IN COOPERATION
IN RELATION TO COMMUNITY PARTICIPATION FACTORS

Factor	Rank in Cooperation			
	Very High		High	
	N	%	N	%
1. Church or church school:				
a. attends regularly	19	5.8	100	30.5
b. does not attend regularly	1	4.2	8	33.2
2. Boy or Girl Scout groups				
a. participates regularly	1	1.3	26	32.5
b. does not participate	19	7.0	82	30.1
3. YMCA or YWCA program:				
a. participates regularly	6	6.1	35	35.7
b. does not participate	14	5.5	73	28.7
4. Ballroom dancing classes:				
a. has participated	6	6.7	34	38.2
b. has not participated	14	5.3	74	28.1

TABLE V

DISTRIBUTION OF STUDENTS RATING VERY HIGH OR HIGH IN COOPERATION
IN RELATION TO TELEVIEWING AND READING FACTORS

Factor	Rank in Cooperation			
	Very High		High	
	N	%	N	%
1. Time spent per week watching TV:				
a. 25 hours or more	6	5.0	43	36.1
b. 20-24 hours	6	8.3	22	30.6
c. 15-19 hours	3	7.5	6	15.0
d. 10-14 hours	5	7.7	18	27.7
e. 5-9 hours	0	0.0	14	42.4
f. 1-4 hours	0	0.0	5	25.0
g. none at all	0	0.0	0	0.0
2. TV preferences:				
a. comedy				
b. westerns	2	3.1	14	21.9
c. musical	8	11.6	22	31.9
d. adventure	2	5.0	18	45.0
	6	4.7	38	29.7
3. Books read during past eight months:				
a. 50 or more				
b. 40-49	6	11.5	19	36.5
c. 30-39	0	0.0	6	28.6
d. 20-29	1	4.2	8	33.3
e. 10-19	7	9.7	19	26.4
f. 1-9	5	5.3	27	28.4
g. none	1	1.3	27	34.2
	0	0.0	2	22.2
4. Preference in book types:				
a. fiction				
b. biography	14	6.0	71	30.3
c. science	4	8.2	16	32.7
d. historical	1	4.0	7	28.0
	0	0.0	9	45.0
5. Number of comic books read per week:				
a. 20 or more				
b. 10-19	1	2.7	11	29.7
c. 5-9	3	5.4	18	32.1
d. 1-4	7	11.5	21	34.4
e. none	8	5.9	43	31.6
	1	1.6	15	24.2
6. Reading achievement level:				
a. 6 months or more above GP				
b. at GP plus or minus 6 months	2	2.3	32	37.2
c. 6 months or more below GP	1	1.5	27	39.7
	1	2.6	7	17.9
7. Change in reading achievement in six month period:				
a. 1 year or more gain				
b. 6-10 months gain	1	1.4	23	31.1
c. 0-5 months gain	1	2.7	14	37.8
d. 1-5 months loss	2	2.9	25	35.7
e. 6 months or more loss	0	0.0	4	36.4
	0	0.0	0	00.0

all or most of their lives; had lived in the community half or more of their lives; and had attended the present school for two or more years. Stability of home and school life had a slightly favorable effect upon children's responses to preferences for cooperation.

Those children who attended church or church school and those who attended a boy or girl scout group regularly did not score significantly higher on cooperation than those who participated less. However, the children who attended YMCA, YWCA, or ballroom dancing classes did rate higher than non-participants in these activities. Of those who did attend YMCA or YWCA activities, the children who participated for five or more hours per week rated considerably higher in cooperation than those of lesser participation.

A student's relative popularity with his classmates seemed to have little if any relationship to his rank on cooperation. Those children who were selected on a sociometric question by five or more classmates did not score significantly differently from those not chosen or those chosen by a lesser number.

Children selected as class officers did not score higher than non-selectees. In this school an outstanding citizen is selected in each classroom each week. Children so selected and those who served on the school council, and/or school safety patrol tended to rate very high or high on cooperation more frequently than those who were not selected for these positions.

A study of television viewing habits indicated that more than half of the children estimated that they watched television more than twenty hours per week. Only three children of the 352 stated that they did not watch television at all. Those who did spend over twenty hours per week in televiewing also tended to rank high in cooperation, and did not appear to have program preferences which differed from those who scored lower.

Children who read extensively also tended to score high or very high on cooperation. About half of the children stated they had read twenty or more books (other than textbooks) during the preceding eight months. Those who read more than fifty books in this period scored much higher in cooperation than did those reading fewer books.

Over half the children stated that they read less than five comic books per week, although over one fourth of the children stated that they read more than ten per week. Those who did read from five to nine comic books per week rated significantly

higher in cooperation than did those who read no comic books.

A comparison of reading achievement with ranks on cooperation indicated that those reading at six months or below their grade placement scored considerably lower than those reading at or above their grade placement. This is understandable as the Behavior Preference Record is a group test requiring the child to be able to read material at his grade level.

Findings and Conclusions

1. Boys tended to score higher in preferences for cooperation than did girls.

2. Stability of home and school life had a slightly favorable effect upon children's preferences for cooperation.

3. Participation in YMCA or YWCA children's activities appeared to have a closer relationship to preferences for cooperation than did participation in other community activities.

4. Preferences for cooperation did not seem to be a factor related to a child's popularity with his classmates except when it came to making choices for the school council and the school safety patrol.

5. Children who watched television twenty hours or more per week expressed greater preferences for cooperative behavior than did those who viewed television a lesser number of hours.

6. Children who read extensively indicated greater preferences for cooperative behavior than those less interested and those less capable in reading.

The Behavior Preference Record is a useful instrument for the teacher interested in studying preferences of his pupils for cooperative behavior. It can be utilized as a discussion guide for encouraging group reactions to problem situations requiring choices of behavior. A comparison of pupil ranks in cooperation with certain factors of home, community, and school life reveals some tendencies of possible value to students of child development. Further study is needed to determine the relationships of children's preferences for cooperative behavior as indicated by this instrument and the actual behavior of the children in group situations requiring group behavior. The influence of cultural expectations and the need for peer approval undoubtedly have a considerable effect upon what the child actually does as contrasted to what he indicates are his preferences for behavior in written test situations.

TABLE VI
DISTRIBUTION OF STUDENTS RATING HIGH OR VERY HIGH
IN COOPERATION IN RELATION TO SCHOOL FACTORS

Factor	Rank in Cooperation			
	Very High		High	
	N	%	N	%
1. Length of enrollment in this school:				
a. all school years	7	4.0	59	33.3
b. three years or more	5	7.3	21	30.9
c. two years	6	10.3	16	27.6
d. one year or less	2	4.1	12	24.5
2. Child's relationship to peers:				
a. selected as class officer	11	5.9	56	30.3
b. never selected as class officer	9	5.4	52	31.1
c. selected as school council representative	7	7.7	29	31.9
d. never selected as school council representative	13	5.0	79	30.3
e. selected to serve on safety patrol	11	13.4	20	24.4
f. not selected to serve on safety patrol	5	7.3	19	27.5
g. selected as outstanding citizen of class	9	4.7	65	33.9
h. not selected as outstanding citizen of class	11	6.9	43	26.9
i. selected as friend:				
1) by five or more peers	2	5.7	12	34.3
2) by four peers	2	8.7	6	26.1
3) by three peers	2	4.2	13	27.1
4) by two peers	4	5.3	24	31.6
5) by one peer	4	4.3	25	26.6
6) not selected	6	7.9	28	36.8

REFERENCES

1. Carmichael, J. (Editor). Manual of Child Psychology, New York. John Wiley and Sons, 1946 pp. 707-751.
2. Heffernan, H. "The Teacher Helps Children to Build Values," California Journal of Elementary Education. Vol. 23: 243-253, May, 1958.
3. Meredith, C. W. "Personality and Social Development During Childhood and Adolescence," Review of Educational Research. Vol. 25: 469-476, December, 1955.
4. Perkins, F. T. "Research Relating to Values," California Journal of Elementary Education. Vol. 23: 223-242, May, 1955.
5. Symposium--"Development of Children's Values," British Journal of Educational Psychology. Vol. 27: 77-88; 28: 1-14, 97-111, June, 1957--June, 1958.
6. Wann, K. and Foshay, A. Children's Social Values. New York, Bureau of Publications, Teachers College, Columbia University, 1954.

MEASURING ATTITUDINAL OUTCOMES OF CHILD PSYCHOLOGY WITH THE PARENTAL ATTITUDE RESEARCH INSTRUMENT

FRANK COSTIN
University of Illinois

IN A STUDY previously carried out by this investigator (1) in which Shoben's (3) scale was used, it was found that students expressed less dominating, possessive, and ignoring attitudes toward parent-child relationships after concluding an undergraduate course in child psychology. The greatest change was in dominating attitudes. Since the publication of this study, other investigators (5, 6) using Shoben's inventory have reported that they found significant changes in attitudes of female home economic students. The present paper reports the results of employing a newer inventory to assess attitude changes--the Parental Attitude Research Instrument (PARI) (2). This instrument was developed on the basis of a factor analysis, and is designed to measure more specific and homogeneous attitudes than the three broad categories represented by Shoben's scale. It would appear, therefore, to have some advantage over Shoben's instrument for measuring attitudinal outcomes in child psychology and related courses.

Procedure

The complete PARI consists of 32 five-to-ten-item scales which measure parental attitudes toward specific aspects of child rearing and family life. The questionnaire used in this study was a selected combination of these scales, totaling 105 items. This briefer form was adopted to meet time limitations of classroom testing. According to Schaefer and Bell's analysis (2), these 105 items measure 5 factors of parental attitudes: suppression and interpersonal distance (35 items), hostile rejection of the homemaking role (30 items), excessive demand for striving (35 items), overpossessiveness (25 items), and harsh punitive control (35 items). Some of these items measure more than one factor.

The PARI was given at the beginning and end of the semester to two undergraduate classes in child psychology, totaling 100 students, and representing a variety of majors and colleges at the University of Illinois. None of these students had ever had a course in child psychology, child development, or family relations, although they all had had an intro-

ductory course in psychology. Course organization and content were the same for each class, with emphasis on parent-child relationships. Course materials and activities included a basic text, lectures, films, discussions, and supplementary readings. The same instructor taught both classes.

The PARI was also administered at the beginning and close of the semester to 54 students enrolled in a social science survey course and to 29 students in an introductory psychology class at the University of Illinois. The reason for testing the social science class was to see what effect general university experience, excluding a course in child psychology, might have on attitudes toward parent-child relationships. None of these students had ever taken a course in child psychology, child development, or family relations, nor were they enrolled in one at the time of the testing. While the social science course did contain some material on the sociology of the family, it did not deal directly or specifically with problems of parent-child relationships. The child psychology and social science classes were approximately the same in the varieties of majors and colleges represented.

The introductory psychology and child psychology classes were taught by the same instructor. The PARI was given to the beginning psychology class to discover whether the instructor's attitudes "as a person" might affect students' attitudes toward parent-child relationships independently of course content. (This possibility was not tested in the investigator's initial study using the Shoben inventory.) The introductory psychology class was very much like the child psychology classes as far as colleges and majors represented, but because it was a lower-division course it contained a higher percentage of freshmen and sophomores. Only a very small portion of this course dealt with parent-child relationships.

Students were asked to respond to the PARI items by strongly agreeing, mildly agreeing, mildly disagreeing, or strongly disagreeing. Each response was scored 1, 2, 3, or 4. The higher the score for a given response, the more intense was the attitude (and therefore less permissive or accepting); the

lower the score, the less intense was the attitude. The total score for each of the five attitude factors was obtained by summing the scores of responses to each item which contributed to that factor.

Results

At the end of their course in child psychology, students revealed significant mean decreases in the intensity of their expressed attitudes. These changes, as shown in Table I, occurred for four of the five factors measured. The greatest mean decrease (7.38) was in "excessive demand for striving", while the next greatest decrease (5.30) was in "harsh punitive control". Smaller mean decreases occurred for attitudes of "suppression and interpersonal distance" (3.53) and "overpossessiveness" (3.22).

In computing the standard error of the difference in means, the correlation between precourse and postcourse scores was taken into consideration. In general, these correlation coefficients (Table I) show that while child psychology students changed significantly in their attitudes, their relative position with respect to these attitudes did not tend to vary markedly.

Neither the social science nor the introductory psychology class showed significant changes in attitude intensity (Table I). Evidently, then, it was the taking of a specific course, child psychology, which led to more accepting and permissive attitudes toward parent-child relationships, since neither general university experience, nor the attitude of the instructor "as a person", independent of course content, were influential in changing attitudes.

Attitude changes of students achieving in the upper half of their class were compared with changes occurring in the lower half. ("Achievement" was determined on the basis of objective examinations covering course content.) Table II summarizes the results of this comparison. While tests of significance showed that the upper half was already less intense in its attitudes at the beginning of the course, both halves decreased in attitude intensity for four of the five factors. Significance tests, however, revealed no differences in how much each half changed.

A comparison of attitude changes in men and women (Table III) showed that while women were consistently less intense in their attitudes at the beginning of the course than were men, both groups decreased significantly in attitude intensity for the same four factors. Furthermore, tests of significance showed no differences in the extent to which each changed.

Discussion

In general, the findings of this study supported and clarified the major conclusions of the investigator's previous report on the effects of child psy-

chology in which Shoben's inventory was used.

1. Both studies indicated that after taking an undergraduate one-semester course in child psychology, students expressed more permissive and accepting attitudes toward parent-child relationships. Furthermore, the present study showed more clearly that it was the nature of the course itself which was influential in changing attitudes rather than the instructor's "personality" as such, since no attitude changes occurred for an introductory psychology class taught by the same instructor.

2. The extent of the changes reflected by the PARI were somewhat greater than those which had been measured with Shoben's inventory. This may indicate that the PARI is a more sensitive instrument.

3. The greatest change revealed by Shoben's questionnaire was in "dominating" attitudes. In the present investigation, the largest changes occurred for two factors which were specific manifestations of that category: "excessive demand for striving", and "harsh punitive control".

4. Both investigations showed that high achievers were already more permissive at the beginning of the course than were low achievers, but that both groups changed equally as much in the direction of permissiveness. Thus, in spite of the fact that some students had attained significantly less information about child psychology than others, they still changed as much, on the average, as did those who showed a greater acquisition of knowledge. These findings lend support to the argument for measuring directly and specifically those attitudes which an instructor hopes to influence. To infer such changes simply on the basis of academic achievement may give the instructor false information as to which students modified their attitudes and which did not.

5. Neither study showed any difference between the extent to which men and women changed their attitudes. In the present investigation, however, women were consistently more permissive than men at the beginning of the course; the Shoben inventory had revealed no such differences. As far as the investigator could determine, the populations of the two studies were not significantly different in terms of general student characteristics. It may be, therefore, that the PARI is better able to measure women's attitudes toward parent-child relationships than Shoben's questionnaire--at least for the kind of population used in this study. (This conclusion is consistent with the findings of older studies showing that women were more "liberal" than men in their attitudes toward children (4)).

Why was there no change in "hostile rejection of the homemaking role"? An examination of the items contributing to this factor revealed that a large proportion of them were concerned with unhappiness resulting from wives being confined to home, fear of routine household duties, fears of children interfering with freedom of movement, and marital conflict. While these kinds of relationships

TABLE I

ATTITUDES OF CHILD PSYCHOLOGY, SOCIAL SCIENCE, AND INTRODUCTORY PSYCHOLOGY STUDENTS TOWARD PARENT-CHILD RELATIONSHIPS

Attitude Measured	Before Course		After Course		r of scores	Difference in means	t
	Mean Score	SD	Mean Score	SD			
Suppression and inter-personal distance							
Child Psychology (N = 100)	62.12	7.56	58.59	7.96	.73	3.53	6.09*
Social Science (N = 54)	63.98	6.81	64.91	7.54	.63	.93	1.10
Introductory Psychology (N = 29)	63.65	6.51	64.00	6.91	.66	.35	.33
Hostile rejection of homemaking role							
Child Psychology	79.41	10.00	80.02	9.89	.81	.61	.98
Social Science	80.54	10.74	81.83	10.62	.84	1.29	1.57
Introductory Psychology	82.03	11.46	83.91	10.39	.90	1.88	2.04
Excessive demand for striving							
Child Psychology	70.87	13.99	63.49	11.66	.78	7.38	8.39*
Social Science	77.30	12.78	76.41	13.13	.82	.89	.85
Introductory Psychology	80.52	14.47	78.34	13.68	.87	2.18	1.60
Over-possessiveness							
Child Psychology	45.17	8.90	41.95	7.96	.73	3.22	5.11*
Social Science	47.28	8.82	46.78	9.44	.82	.50	.67
Introductory Psychology	48.28	10.13	47.17	11.00	.90	1.11	1.24
Harsh punitive control							
Child Psychology	73.01	13.07	67.71	12.33	.81	5.30	6.71*
Social Science	77.09	13.30	77.17	13.93	.84	.08	.07
Introductory Psychology	79.14	15.01	78.76	13.73	.87	.38	.27

* $p < .01$

Note -- The higher the score, the more intense the attitude.

TABLE II
ATTITUDES TOWARD PARENT-CHILD RELATIONSHIPS OF STUDENTS
ACHIEVING IN UPPER AND LOWER HALVES OF CHILD PSYCHOLOGY

Attitude Measured	Before Course		After Course		r of scores	Difference in means	t
	Mean Score	SD	Mean Score	SD			
Suppression and inter-personal distance							
Upper half (N = 50)	60.26	7.40	56.48	8.03	.81	3.78	5.56*
Lower half (N = 50)	63.98	7.26	60.70	7.29	.60	3.28	3.57*
Hostile rejection of homemaking role							
Upper half	78.20	8.87	77.86	10.08	.81	.34	.40
Lower half	82.18	10.88	80.62	9.21	.83	1.56	1.79
Excessive demand for striving							
Upper half	66.24	13.43	58.26	11.04	.72	7.98	5.96*
Lower half	75.50	12.98	68.72	9.72	.78	6.78	5.85*
Over-possessiveness							
Upper half	42.98	8.58	38.96	7.52	.74	4.02	4.85*
Lower half	47.36	8.67	44.94	7.23	.68	2.42	2.62*
Harsh punitive control							
Upper half	67.80	12.27	62.68	12.20	.82	5.12	4.89*
Lower half	78.22	11.71	72.74	10.22	.72	5.48	4.67*

Note -- The higher the score, the more intense the attitude.

* $p < .01$

TABLE III
ATTITUDES TOWARD PARENT-CHILD RELATIONSHIPS OF MEN
AND WOMEN STUDENTS IN CHILD PSYCHOLOGY

Attitude Measured	Before Course		After Course		r of scores	Difference in means	t
	Mean Score	SD	Mean Score	SD			
Suppression and inter-personal distance							
Men (N = 47)	64.11	8.07	60.21	7.74	.74	3.90	4.70*
Women (N = 53)	60.36	6.55	57.15	7.90	.70	3.21	4.06*
Hostile rejection of homemaking role							
Men	82.59	8.31	81.57	8.61	.74	1.02	1.15
Women	77.34	10.98	77.49	10.40	.83	.25	.28
Excessive demand for striving							
Men	73.91	13.54	65.34	11.06	.79	8.57	7.14*
Women	68.17	13.85	61.85	11.93	.75	6.32	4.96*
Over-possessiveness							
Men	47.94	9.15	43.83	8.21	.73	4.11	4.40*
Women	42.72	7.89	40.28	7.35	.69	2.44	2.94*
Harsh punitive control							
Men	76.51	12.61	70.91	10.81	.81	5.60	5.17*
Women	69.91	12.69	64.87	12.89	.79	5.04	4.46*

Note -- The higher the score, the more intense the attitude.

* $p < .01$

were not completely ignored in the course, they were dealt with relatively little as compared with other kinds of attitudes measured by the PARI. It is also of interest to note that this attitude factor was already more intensely expressed at the beginning of the course than any of the other four factors. It may be that the factor tapped more readily some rather conscious apprehensions commonly occurring in college students--concerns that marriage will have a restraining influence on their behavior. Furthermore, this is the kind of attitude which is likely to be reinforced by these students' recent (and for some, current) attempts to win psychological independence from parental control.

In conclusion, this study shows that the PARI is a useful device for measuring attitudinal outcomes of child psychology, or related kinds of courses. The results also suggest the importance of undertaking other investigations such as these: effects on student attitudes of different approaches to organizing and teaching child psychology; how child psychology students perceive their instructors' attitudes toward parent-child relationships, and the correlation of these perceptions with changes in their own attitudes; correlations between child psychology students' attitudes and their parents' attitudes toward family relationships. These are all areas of research which could be explored with the PARI.

Summary

The Parental Attitude Research Instrument (PARI) was administered at the beginning and end of the semester to two undergraduate classes in child psychology. Postcourse tests showed significant decreases in attitude intensity for four factors: "excessive demand for striving", "harsh punitive control", "suppression and interpersonal distance", and "over-possessiveness". The greatest changes were in the first two factors named. No changes occurred in the fifth factor measured, "hostile rejection of the homemaking role". Two control groups who were given the PARI before and after the semester showed no changes in attitudes. They consisted of a social science survey class of 54 students, and an introductory psychology class of 29

students. The latter class was taught by the instructor of the child psychology classes.

Students achieving in the upper half of the child psychology classes changed as much as those in the lower half, although upper half students were more permissive in their attitudes at the beginning of the course. No sex differences in the amount of attitude change appeared; women students, however, were already more permissive at the beginning of the course than were men.

In general, the findings supported the investigator's previous study in which Shoben's parental attitude scale was used. The present findings, however, delineated greater and more specific changes in the direction of permissive and accepting attitudes.

The PARI appears to be a promising device for carrying out a variety of research studies on the teaching of child psychology.

REFERENCES

1. Costin, Frank. "The Effect of Child Psychology on Attitudes Toward Parent-Child Relationships," *Journal of Educational Psychology*, 1958, 49, 37-42.
2. Schaefer, S. and Bell, R. Q. *Parental Attitude Research Instrument (PARI)*, Bethesda, Md.: National Institute of Mental Health, Child Development Section, no date (Mimeographed).
3. Shoben, E. F., Jr. "The Assessment of Parental Attitudes in Relation to Child Adjustment," *Genetic Psychology Monographs*, 1949, 39, 101-148.
4. Stogdill, R. M. "Experiments in the Measurement of Attitudes Toward Children: 1890-1935," *Child Development*, 1936, 7, 31-36.
5. Walters, J. "The Effects of an Introductory Course in Child Development on the Attitudes of College Women Toward Child Guidance," *Journal of Experimental Education*, 1959, 27, 311-321.
6. Walters, J. and Fisher, Clara, "Changes in the Attitudes of Young Women Toward Child Guidance over a Two-year Period," *Journal of Educational Research*, 1958, 52, 115-118.

A COMPARISON OF THE PROBLEMS OF CERTAIN ANGLO- AND LATIN-AMERICAN JUNIOR HIGH SCHOOL STUDENTS

PAUL WITHERSPOON

Arkansas State Teachers College, Conway, Arkansas

SEVERAL STATES have a large Latin-American population. One function of the school is to help its students to adjust better to their problems. The question arises whether the two major racial groups, i. e., Anglo-Americans and Latin-Americans, who attend non-segregated schools, have the same problems, or do they differ in type and acuteness.

For this study, a total of 309 students filled out the Junior High School form of the Mooney Problem Check List: sixty-three Anglo-American girls, ninety-seven Latin-American girls, seventy-one Anglo-American boys, and seventy-eight Latin-American boys, in grades seven and eight of a junior high school in San Antonio, Texas. The school is located in a combination business-residential section of town and has about an equal number of Latin and Anglo-Americans. The neighborhood as a whole is composed of a somewhat below average socioeconomic stratum.

The Mooney Problem Check List, Junior High School Form, 1950 Revision, consists of 210 problems divided into seven areas with thirty items in each area. The seven areas are: (1) Health and Physical Development (HPD), (2) School (S), (3) Home and Family (HF), (4) Money, Work, the Future (MWF), (5) Boy and Girl Relationship (BG), (6) Relations to People in General (PG), and (7) Self-Centered Concerns (SC).

The students were instructed to underline the problems troubling them, and then to circle the problems troubling them most. The usual precautions were taken to get valid responses.

The two groups are compared under four groupings, namely, Latin girls, Anglo girls, Latin boys, and Anglo boys. Only those problems marked by at least 20 percent of one of the four groups were considered in this study; and only those showing at least a 10 percent difference will be discussed. The problems of most concern are shown if at least 10 percent of one of the groups marked it.

Out of eighteen school problems, the Latin-Americans have a higher percentage in twelve. Five of the problems are common to the boys and girls. Sixteen percent more of the Latin boys and 20 per-

cent more of the Latin girls are afraid of failing in school work than are the Anglos. Both of the two races have trouble with arithmetic, but a higher percentage of the Latins have trouble here. Feeling that they are not smart enough is indicated by 35.9 percent of the Latin boys and 41.1 percent of the Latin girls, as compared to only 11.2 percent of the Anglo boys and 11.1 percent of the Anglo girls. Also, more Latin boys and girls are afraid to speak up in class than are Anglo boys and girls. About 25 percent of the Anglo boys and 35 percent of the Anglo girls feel that the teachers do not practice what they preach, as compared to only 9.8 percent of the Latin boys and 9.2 percent of the Latin girls.

Latin boys have more school problems than Latin girls, and to a greater degree than Anglo boys, in getting low grades, in spelling and grammar, and in writing. The Latin boys also have more trouble in reading and making oral reports. The Latin girls worry more than Anglo girls about their grades.

The Anglo boys and girls do have some school problems to a greater degree than do the Latins. About 25 percent of the Anglo boys feel that school is too strict as compared to only 15.4 percent of the Latin boys. Also, 23.8 percent of the Anglo girls feel there is too little freedom in classes as compared to only 10 percent of the Latin girls. A greater percentage of the Anglo girls are not interested in books and are bored with dull classes than are the Latin girls.

Tables II and III show the problems troubling students most. Eight out of the twelve problems troubling the Latin girls most are school problems. Only one of the seven problems troubling Anglo girls most is a school problem, and that is dull classes. The Latin girls' greatest problem is that of getting low grades in school. They also worry about arithmetic, fear to fail in school work, and feel that they are not smart enough. Four of the eight problems troubling the Latin boys the most are in the school area and indicate fears of failure and worry about grades. The Anglo boys have four

TABLE I

SCHOOL PROBLEMS SHOWING AT LEAST A 10 PERCENT DIFFERENCE BETWEEN
ANGLO- AND LATIN-AMERICAN BOYS AND GIRLS

Problem	Percent of Boys		Percent of Girls	
	Anglo	Latin	Anglo	Latin
Getting low grades in school	35.2	62.8	23.8	43.3
Afraid of failing in school work	30.0	46.0	34.9	46.4
Trouble with arithmetic	33.8	59.0		
Trouble with spelling or grammar	20.0	32.0		
Not spending enough time in study	20.0	35.9		
Can't keep my mind on my studies	28.2	42.3		
Trouble with oral reports	20.0	32.2		
Not interested in certain subjects	38.0	28.2		
School is too strict	25.3	15.4		
Teachers not practicing what they preach	25.3	9.8	34.9	9.2
Slow in reading	15.5	29.5		
Trouble with writing	10.0	20.0		
Not smart enough	11.2	35.9	11.1	41.1
Afraid to speak up in class	15.5	32.0	20.6	49.5
Worried about grades			28.6	47.4
Not interested in books			28.6	6.2
Too little freedom in classes			23.8	10.0
Dull classes			20.6	11.4

problems in the school area, but not to the extent the Latins do.

School does present problems to both races, and it is up to the schools to help their students solve their problems. The Latin-Americans have more school-connected problems than do the Anglo-Americans, and since it is possible to identify many of the specific problems troubling them, it should be easier for the schools to help solve these problems.

The problems in the other six areas are not directly school problems, but should be of interest to counselors and others interested in helping students. The results will be summarized as briefly as possible.

In the Money, Work and Future area, Latin girls again have more problems and to a greater degree. Of the six problems in this area showing at least a 10 percent difference between the girls, the Latin girls have five to a greater degree. These are deciding what to take in high school, needing a job during vacations, having no regular allowance, having to ask parents for money, and not knowing how to look for a job. More of the Anglo girls had as a problem wondering if they would ever get married. The Anglo boys had one more problem in this area than did the Latin boys and that was wanting to know about trades.

The Anglo girls had more problems in the Self-centered Concerns area. More Anglo girls than Latin girls are bothered by being afraid of making mistakes, not being able to forget some mistakes they have made, and daydreaming too much. The Latin girls are more concerned over forgetting things. A higher percentage of Latin boys than Anglos have the problems of lacking self-control, being lazy, and not taking some things seriously enough.

In the Boy and Girl Relations area, more Latin girls have the problems of learning how to dance and dating more frequently than do the Anglo girls. The boys do not show much difference here, but as problems of most concern, the Anglo boys mark not being allowed to run around with kids they like, and too little chance to do the things they want to do. The Latin boys mark no problems to a significant degree in this area.

In the Home and Family area, the girls seem to have about the same frequency of problems. As

problems troubling them most, however, one out of every ten Anglo girls marked the problems of parents not understanding them and talking back to their parents, while the Latin girls marked no problems above the 10 percent level of frequency in this area. The Anglo boys gave as a problem that they wanted things their parents would not give them, but the Latin boys had no significant problems here.

In the Relations to People in General area, the Latin girls are more frequently concerned over being jealous, and the Anglo girls are more concerned with wanting a more pleasing personality. The Anglo boys are more concerned over being bashful and wanting a more pleasing personality. The Latin boys do not show any significant problems in this area.

The Health and Physical Development area has one problem in it bothering at least 10 percent more of the Latin girls than Anglo girls, and that is they often are not hungry for their meals. However, as a problem bothering them most in this area, 10.3 percent of the Latin girls often have headaches and feel that they are not good-looking. The Anglo girls marked no significant percentage of problems in this area. In this area many of the Latin boys are worried about being underweight, and this is also a problem bothering them most. As a matter of most concern to the Anglo boys they mark not getting enough sleep.

It is not the purpose of this study to attempt to point out how the school can help its students solve their problems, as that can be done only by experts in the field. It is hoped that school counselors and administrators, by knowing what the main problems of the Anglo-American and Latin-American students are, can mobilize their counseling and curricular offerings to meet these problems. A program could be planned to meet these indicated problems through curricular and extra-curricular means, taking into account some of the differences between Anglo and Latin-American students. There are more likenesses than differences in the problems of the two groups, and these possibly should be dealt with first, but special attention could be given to the more frequent and acute school problems of the Latin-Americans.

EFFICACY OF TWO TESTS IN DIFFERENTIATING POTENTIALLY LOW FROM AVERAGE AND HIGH FIRST GRADE ACHIEVERS

ELMER F. MORGAN JR.
Dundee, Michigan

Problem

The problem of the present study was to discover whether the Pintner-Cunningham Primary Test, Form A (13) and the Full-Range Picture Vocabulary Test, Form B (1, 2, 3, 4) have predictive value at the late kindergarten and early first grade level.

The basic assumption made here is that IQ (mental age) is the best single predictor of potential school achievement (6, 10). Relevant to this assumption is the work of Raybold (14) and Woods (15), who have reported that a mental age of 76 months was adopted by several large California school systems as the best dividing line between potential success or failure in beginning reading. Children with mental ages below this figure (approximately 50%) were enrolled in special "transition B1" classes. Betts (5) has noted that such a plan is not financially feasible for many schools and has discussed the pros and cons of grouping for reading instruction within the class. Still other schools are experimenting with the plan of homogeneous grouping at the early elementary level.

Crucial to such a plan, however, is whether or not a valid differentiation concerning levels of potential achievement can be made at the late kindergarten level. This differentiating process usually becomes the responsibility of kindergarten teachers who are not trained especially for this work. This study will try to ascertain: (a) whether specially trained kindergarten teachers can make valid potential achievement level placements, and (b) whether or not the above named tests would prove of value in making such decisions.

The major task of this evaluation was to determine the empirical validity of the two tests by: (a) assessing their parallel validity; (b) establishing predictive validity to an outside criterion; and (c) discovering whether the above validities included an identifiable relationship between test scores and a criterion measurement of correct grouping and first grade teachers.

Description of Tests

The Pintner-Cunningham Primary Test (P-C) is a group test published in three equivalent forms. The test is composed entirely of pictures and is marked according to examiner's verbal instructions. The test manual (13) reports reliability coefficients of .88 for 60 kindergarten and .94 for 216 students, kindergarten through grade two. Validity coefficients between the P-C and Stanford Revision of the Binet Scale are high. Daugherty (7) reports an r of .80 for a kindergarten group and Pintner (12) reports r 's of .73 and .88 for 229 and 72 cases. In two separate studies Grant (8, 9) found an r of .63 between reading test scores and an r of .81 between total scores of the Metropolitan Readiness Tests. Thus, in terms of parallel reliability and validity the P-C seems to be satisfactory for most school purposes. The author, however, knows of no studies concerning the validation of the P-C to an outside criterion such as teacher prediction concerning potential achievement.

The Full-Range Picture Vocabulary Test (FRPVT) is an easy to administer individual test consisting of 16 plates, 8 1/2 x 11 inches, each with four cartoon-like drawings. The drawings are of common objects, familiar scenes, and human activities. There are parallel forms consisting of eighty-four stimulus words per form. For examination purposes the examiner keeps the answer sheet of stimulus words and asks testee which one of the four drawings best represents a particular stimulus word. The testee can respond by pointing or by any method indicating a choice of yes or no. Thus, the test can be given to anyone capable of hearing and capable of signalling yes or no. A mental age and IQ can be obtained within ten minutes at the primary level.

A study by Ammons and Holmes (3) appears to be the only study in the literature concerning the use of the FRPVT with a preschool-age population. Their population consisted of 120 children. From the chronological age levels, two through five, 15 boys and 15 girls controlled for age, sex, race and

occupational status were tested. Only native-born white children were included, and all were from the Denver area. The results from this population constitute the norms for the above age levels. Reliability coefficients between parallel forms were .76, .79, .83, .78, and .93 for the age levels 2, 3, 4, 5 and the whole group respectively. Stanford-Binet (1937) mental ages and FRPVT scores correlated .85 for Form A and .83 for Form B over the whole age range, indicating a satisfactory validity.

The reason for selecting the FRPVT was its ease of administration. It was thought that a qualified school psychologist could train a competent counselor or teacher to administer the test effectively, should it prove to be a valuable supplement to a group testing program.

Subjects and Procedure

The Pintner-Cunningham Primary Test, Form A (P-C) was administered to 80 kindergarten pupils in late May of 1958. This group of subjects represented all of the kindergarten pupils from a rural consolidated school (school population 964). This school is closely surrounded by rich farms and several cities ranging in population of from 20,000 to 300,000. Past testing throughout the various grades tends to suggest that the norms for this school are somewhat above the national level. Thus results from this study will be suggestive rather than representative of other school districts.

The P-C tests were not scored, due to a complete change of the school administration, until late summer. Thus, the three kindergarten teachers made their recommendations for individual group placement completely independent of test scores. All recommendations were followed. At the start of the fall term 76 students remained. Sixteen had been placed into the potentially low achievement group, 28 into the potentially average achievement group. Two weeks after the first six weeks marking periods several tasks were performed: (a) The FRPVT was administered to the entire low group, 15 randomly selected students of the high group, and an overall random sample of 50 students; (b) P-C test score predictions and actual group placements by kindergarten teachers were compared by appropriate statistical techniques; (c) first grade teachers were asked to make a list of all the children in their classes who appeared to be either too advanced or too slow for present class placement (disagreement with kindergarten teacher placement); and (d) first grade teachers were asked to rank their children in terms of relative brightness, using grade books, subjective hunches, observational data and any other data deemed pertinent. It should be noted that all early elementary teachers in this school have had intensive in-service training concerning maturational growth, individual differences, symptoms of retardation, brightness, etc. The kindergarten teachers have had in-the-class training by a team of university experts. It was felt that these teachers

would be able to make reasonably competent observations and recommendations.

Hypotheses

Previous to examining the P-C scores it was hypothesized that:

1. There would be significant rho's between the P-C and FRPVT IQ's for the extremes (high vs low achievement groups) and that there would be a significant Pearson r between the two tests for the randomly selected sample of 50 students.
2. There would be significant differences between the means of the three homogeneous achievement groups in accord with an outside criterion (kindergarten teacher observations) and that there would not be a statistically significant difference between test score and kindergarten teacher predictions of potential achievement within each separate group.
3. There would be significant rho's between the first grade teachers' perceptions of relative brightness and P-C test rankings and also significant rho's between teacher perceptions of relative brightness and FRPVT test rankings for the low group and 15 randomly selected students from the high group.
4. Those children deemed either too advanced or too slow for present first grade group placement would have P-C scores confirming these observations.

Results

1. Hypothesis 1 was confirmed. Between the P-C and FRPVT IQ's rho's of .87 and .80 were obtained for the slow achievement group and 15 randomly selected students of the high achievement group (all rho's in this study were significant at beyond the .02 level of confidence). A Pearson r of .797 was obtained between the FRPVT and the P-C for the randomly selected sample of 50 students.

2. Hypothesis 2 was confirmed. The means and standard deviations for the slow, average, and high groups were 92.68 and 9.06; 107.36 and 10.36; 113.59 and 10.66 respectively. The t test was significant at less than the .05 level for any two means. Table I shows the number of students in each group based on kindergarten teachers' recommendations and their corresponding P-C test scores. The asterisks represent those children judged to be either too advanced or too slow by present first grade teachers. In the low group it is interesting to note that 13 of 16 test scores were below 100 and that these scores do agree with the outside criterion of kindergarten teacher placements. The three \bar{S} s above 100 were listed as those who appeared to be too fast for their present group placement. That 13 of 16 test predictions agreed with the outside criterion could happen only eleven times out of a thousand by chance alone (expansion of binomial). Two stu-

TABLE I

FREQUENCY DISTRIBUTION OF THE THREE ACHIEVEMENT GROUPS
BASED ON RECOMMENDATIONS OF KINDERGARTEN TEACHERS

P-C Test Scores	Slow Achievers	Average Achievers	High Achievers
140-135	Sample M = 107.26 S. D. = 12.91		
135-130			
130-125		2*	4
125-120		2*	4
120-115	1*	2	7
115-110		4	6
110-105	1*	8	3
105-100	1*	7	4
100- 95	3	0	1*
95- 90	3	2*	1*
90- 85	3	1*	
85- 80	3		
Below 75			

* The asterisks represent those children judged to be either too advanced or too slow for present class placement by present first grade teachers.

dents above 125 in the average group had originally been recommended for a high group placement and cannot be considered true misses. Results for the average and high group are statistically comparable to those described in the slow group.

3. Hypothesis 3, that there would be significant rho's between P-C, FRPVT and teacher rankings of relative brightness, was also confirmed. Rho's of .644, .710, .655 were obtained between P-C test and teacher rankings for the slow, average, and high groups respectively. Rho's of .610 and .642 were obtained between teacher and FRPVT rankings for the slow group and fifteen randomly selected students of the high group.

4. Because of the overlap of test scores between the average and high groups it is somewhat doubtful that a clear-cut confirmation of hypothesis 4 is in order. However, it is somewhat interesting to note that the P-C would have tended to suggest that the 12 misplaced students according to the first grade teachers were actually misplaced. Such an event could happen only one time out of 4,096 by chance alone (expansion of binomial).

5. Several additional tests were applied with some rather surprising results. Validation studies with the Stanford-Binet Scale for both tests suggested that there should be significant correlations, but it seemed unlikely that the means for the two tests would be from a common population. The Wilcoxon Matched-Pairs Signed-Ranks test revealed that there was not a significant difference between the P-C and FRPVT means for either the high or low group ($P > .10$). The mean for the random sample of 50 students was 106.88 for the P-C and 106.91 for the FRPVT. Inspection alone reveals that the means are from a common normal population.

Discussion of Results

The results of this study suggest that: (a) well trained kindergarten teachers can make reasonably competent achievement placements; (b) the above tests would have appreciably improved achievement prediction; (c) the two tests show a statistically significant relationship with the outside criterion of kindergarten teacher placement recommendations; and (d) both tests correlate satisfactorily concerning expected achievement as indicated by grades and observations of first grade teachers.

These findings strongly suggest that both tests have good empirical validity with respect to differentiating the potentially low from the potentially average and high first grade achievers at the late kindergarten level. It is suggested that the FRPVT should prove to be a valuable supplement to a group testing program, especially when group test results and teacher ratings are in serious disagreement. The results of this study further appear to be in line with Meehl's (11) argument that actuarial methods frequently have more predictive efficiency than time consuming staff conferences, no matter how well

trained the participants.

REFERENCES

1. Ammons, R. B., Arnold, P. R., and Hermann, R. S., "The Full-Range Picture Vocabulary Test: IV. Results for a White School Population," *Journal of Clinical Psychology*, 1950, 6, 164-169.
2. Ammons, R. B., and Holmes, J. C., "The Full-Range Picture Vocabulary Test: III. Results for a preschool-age population," *Child Development*, 1949, 20, 5-14.
3. Ammons, R. B., and Huth, R. W., "The Full-Range Picture Vocabulary Test: I Preliminary scale," *Journal of Psychology*, 1949, 28, 51-64.
4. Ammons, R. B., and Rachiele, L. D., "The Full-Range Picture Vocabulary Test: II. Selection of items for final scales," *Educational Psychological Measurement*, 1950, 10, 307-319.
5. Betts, E. A., "Developing Basic Skills . . . Through Effective Class Organization," *Education*, 1958, 78, 1-16.
6. Blommers, P., and Stroud, J. B., "Note on the Organismic Age Concept," *Journal of Educational Psychology*, 1958, 49, 106-107.
7. Daugherty, M. L., *A Comparative Study of Nine Group Tests of Intelligence for Primary Grades*. John Hopkins Press, Baltimore.
8. Grant, Albert, "The Comparative Validity of the Metropolitan Readiness Tests and the Pintner-Cunningham Primary Mental Test," *Elementary School Journal*, XXXVIII, 8, 599-605.
9. Grant, Albert, "A Comparison of the Metropolitan Readiness Tests and the Pintner-Cunningham Primary Mental Test," *Elementary School Journal*, XXXVIII, (Oct. 1937), 118-126.
10. Klausmeier, H. J., Beeman, Alan, and Lehmann, Irvin J., "Comparison of Organismic Age and Regression Equations in Predicting Achievements in Elementary School," *Journal of Educational Psychology*, 1958, 49, 182-186.
11. Meehl, Paul E., *Clinical versus Statistical Prediction*. University of Minnesota Press, Minneapolis.
12. Pintner, R., "The Pintner-Cunningham Primary Test," *Journal of Educational Psychology*, 18, 52-58.
13. Pintner, R., Cunningham, Bess V., and Durst, W. N., *Pintner-Cunningham Primary Test: Manual of Directions*. Yonkers-on-Hudson, N. Y.: World Book Company, 1941.
14. Raybold, Emma, "Reading Readiness in Children Entering First Grade," *Third Yearbook*

<p><u>of the Psychology and Educational Research Division, School Publication No. 185, Los Angeles, California, 1929, 98-101.</u></p> <p>15. Woods, Elizabeth L., and Staff, "A Study of</p>	<p>the Entering B 1 Children in the Los Angeles City Schools," <u>Journal of Educational Research</u>, XXXI (September, 1937), 9-19.</p>
--	--

PUBLICATION ACTIVITIES OF MEMBERS OF THE DIVISION OF SCHOOL PSYCHOLOGY OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION

REGINALD LANIER JONES
Ohio State University

DISCUSSION BY participants of the Thayer Conference (1) indicated that there was difficulty in carrying out research in the schools, although there was general agreement that it should be undertaken. There were questions as to how much research the school psychologist could carry on. Cutts writes (1, p. 45): "The school psychologist appreciated the need for research but also saw the overwhelming need for direct service, which tends to absorb all of the available time. Many doubted that their school systems would permit them to take from other duties the time which research entails, and most of the interviews with the superintendents bear them out."

In spite of the observation that research by school psychologists is desirable but difficult, it is known that many of them are involved in research. To date, little systematic inquiry into their research activities has been reported. Such an investigation would seem to give some indication as to the amount and kind of research which school psychologists are undertaking. In addition, clues as to the direction which the area of school psychology seems to be taking (as indicated by the range of research problems to which the psychologists have addressed themselves) may be obtained.

To accomplish the task of inquiring into the nature of the school psychologists research activities, the publications of members of Division 16 as reported in the *Psychological Abstracts* for the years 1952-1956 were summarized. Analyses were made in terms of the sub-area to which the article applied (i. e., mental deficiency, special education, school learning, etc.), type of publication (i. e., journal article, book, pamphlet, etc.), place of employment of the author (i. e., whether employed by Boards of Education or not), status within the Division (i. e., associate or fellow), location of the author (in terms of town size), and the highest degree held.

Results

Two hundred forty-nine publications written by 92 members of Division 16 were reported in the *Psychological Abstracts* during the five year period

studied. Forty-four percent of these publications were accounted for by fellows who constitute 24 percent of the Division and 56 percent were contributed by associates who constitute 76 percent of the Division. When the two groups were compared statistically, a X^2 of 54.08 was obtained, which is significant beyond the .001 level of confidence.

One hundred ninety-five of the 249 publications were journal articles (reported in 64 different journals). These constituted 79 and 78 percent of the total number of publications for the non-Board of Education employees and the Board employees respectively. The numbers of books authored favors the non-Board group, constituting 11 percent of the total number of publications as compared to 4 percent for the Board group. Data on bulletins, pamphlets, chapters in books, dissertations, theses, tests and test manuals are presented in Table I. Because of the small N in these categories, analysis was not undertaken.

Thirty-six employees of Boards of Education contributed 76 publications while 56 non-board of Education employees accounted for 173 publications. The mean number of publications for the Board of Education employees was 2.17 with a standard deviation of 2.02. The non-Board group had a publication mean of 3.09 with a standard deviation of 2.78. When the two groups were compared, a T score of 1.84 was obtained. This is significant at greater than the .05 level of confidence but less than the .10 level.

The 76 publications of Board employees covered 21 areas (i. e. divisions such as child guidance, diagnosis and evaluation, etc.). The heaviest concentration (55%) was in the areas of diagnosis and evaluation, special education, mental deficiency and educational guidance, each area having 14, 13, 8 and 7 publications respectively. The 173 publications by non-Board of Education employees covered 31 areas. The areas of childhood and adolescence, educational staff personnel, school learning, and special education accounted for 47 percent of the publications, each having 23, 10, 24, and 24 publications respectively. Special education was the only area of overlap in the two groups. This informa-

TABLE I
MODE OF DISSEMINATION OF PUBLICATIONS

Type of Publication	Board of Education Employees (N - 36)	Non-Board of Education Employees (N - 56)	Total
Journal Articles	60	135	195
Books	3	19	22
Bulletins or Pamphlets	0	4	4
Chapters in Books	4	7	11
Dissertations or Theses	8	4	12
Tests	0	4	4
Test Manuals	<u>1</u>	<u>0</u>	<u>1</u>
Totals	76	173	249

TABLE II
AREAS OF PUBLICATION

Publication Area	Bd. of Ed.	Others	Total
Child Guidance	2	3	5
Childhood and Adolescence	2	23	25
Clinical Neurology	0	2	2
Clinical Psychology, Guidance, Counseling	1	2	3
Complex Processes and Organizations	1	1	2
Crime and Delinquency	2	3	5
Cultures and Cultural Relations	0	2	2
Developmental Psychology	0	1	1
Diagnosis and Evaluation	14	9	23
Educational Guidance	7	7	14
Educational Measurement	4	9	13
Educational Psychology	1	4	5
Educational Staff Personnel	3	10	13
History and Biography	0	2	2
Intelligence	1	4	5
Interests, Attitudes, and Habits	4	0	4
Learning and Memory	0	2	2
Mental Deficiency	8	8	16
Methodology, Techniques	1	1	2
New Tests	0	4	4
Personality	2	3	5
Physically Handicapped	1	9	10
Professional Problems	1	6	7
Professions	1	0	7
Psychosis	0	1	1
School Learning	4	24	28
Special Education	13	24	37
Speech Disorders	3	3	6
Statistics	0	1	1
Treatment Methods	0	4	4
Vocational Guidance	0	1	1

tion is presented in Table II.

Twenty-seven or approximately 67 percent of the Board writers held the Ed. D. or Ph. D. They accounted for 58 or 76 percent of the publications. The remaining 18 publications were accounted for by 9 psychologists holding the master's degree. For the non-board authors, it was found that 92 held the doctorate and 8 percent held the master's degree. The former group accounted for 94 percent of the publications while the latter group accounted for the remaining 6 percent.

When an analysis of the locations of the board writers is made, one finds that 11 authors employed on a county, area or state basis accounted for 25 publications; 21 authors working in cities having populations of 150,000 or over accounted for 41 publications; and 4 authors working in cities under 50,000 (actual range from 10,000 to 32,000) contributed the remaining 5 publications. When those authors employed on a county, area or state basis are combined with those employed in cities having populations over 150,000, 89 percent of the authorship and 93 percent of the total number of publications are accounted for.

Discussion

The results indicated that fellows, although proportionately smaller in numbers, were greater contributors to the literature. The greater experience and education of this group are factors which are probably related to their increased productivity.

A number of distinctions were found between the group employed by Boards of Education as opposed to those not so employed. The non-board group (composed largely of university-connected personnel but supplemented by individuals in private practice and in institutions) was more productive from a research point of view. Related factors here are similar to those present in the "fellow" group. It appears highly likely that this group has more time and education, and greater research skills than the board-connected group. These same factors probably account for the increased number of books which the non-board group has authored.

Inspection of the areas of greatest productivity of board and non-board employees suggests that the interests and/or activities of the two groups varies. The board group's publications seemed more related to the immediate work problems of the school psychologist, while those of the non-board group seemed (based on inspection of areas of maximum productivity) less related to the immediate problems of the school psychologist, though certainly

related to his functions as broadly conceived.

For the board group the findings that individuals holding the doctorate were more active from a research point of view is probably one which would have been predicted. Also noteworthy in the line of expectations is the finding that individuals employed in the larger cities (population of 150,000 or more), and those employed on a county, area or state level, are more productive from a research point of view. The variables of time, training and experience are probably related to the increased productivity of these groups.

Findings

The following statements are drawn from an analysis of the data.

1. Individuals in Division 16 have addressed themselves to a large number of topics, all of which are related to the area of school psychology as broadly conceived.

2. The areas of emphases by those writers employed by Boards of Education seem to be somewhat different from those not so connected, although some overlap exists. The publications of the board group seems related more closely to immediate problems which the school psychologist is likely to face.

3. Individuals in the Division not connected with Boards of Education tend to be greater contributors to the literature. The greater amount of education and training of this group in addition to their probably more adequate research skills and time available for research are suggested as factors accounting for their greater productivity.

4. Among the board group, those holding the doctorate were more productive.

5. Those board employees in the larger cities and those employed on a county, area, or state level were more productive (accounting for 93 percent of the publications for this group).

6. Although board employees constitute 60 percent of the Division, 69 percent of the publications were written by those not directly connected with the schools.

7. Fellows are more productive than associates from the point of view of the number of publications.

REFERENCE

1. Cutts, Norma E., editor. School Psychologists at Mid-Century. Washington: American Psychological Association, 1955.

CHILDREN'S PREFERENCES IN PICTURE STORY BOOK VARIABLES

RUTH HELEN AMSDEN
Cornell University, Ithaca, New York

MORE BOOKS WITH more illustrations than ever before are currently made available for the use of the preschool child. Educators and parents, as well as authors, illustrators, and editors ask the question--what books appeal to the young child? Books presented to the three to five year old child may have a lasting influence upon his attitudes toward books and reading. The investigation reported here was designed to discover three to five year old children's preferences in illustrations and story subjects. The present article concerns the research and conclusions involving the illustration preferences.

The Problem

Many factors within an illustration influence its appeal. The subject matter, artistic quality, style of drawing, arrangement of colors, amount of color, value of color and the mood in a picture all effect the viewer. A few of these variables were isolated in this study to help answer the question: What amount of color, value of color, and kind of drawing do three, four and five year old children prefer in illustrations?

The Children in the Study

Sixty boys and girls, aged three to five, from three different nursery schools composed the research population. There were twenty children in each age group, ten girls and ten boys. Two nursery schools were located in Ithaca, New York and the third in Woodstock, Vermont.

Materials and Procedure

Two sets of ten illustrations were prepared with identical subject content for all the pictures within a set. Each picture within a set varied in amount of color, value of color or style of drawing. Two basic black and white photographs (from a picture story book) acted as the standard subjects for the two sets. Line drawings were made from the photographs and painted according to the variable to be

tested. The two sets were thus composed:

To observe the amount of color preferred--

1. A black and white line drawing.
2. A drawing with one saturated color.
3. A drawing with two saturated colors.
4. A drawing with three saturated colors.
5. A drawing with four saturated colors which also represented true-to-life style and full color value.

To observe the value of color preferred--

6. A drawing with lighter tints of the four colors.
7. A drawing with darker shades of the four colors.

To observe the preference for photographs or line drawings--

8. A black and white photograph.

To observe the style of drawing preferred--

9. A modified realistic drawing of four saturated colors (few dark lines, omitting facial features.)
10. A fanciful drawing of four saturated colors (fanciful element of personified dish and spoon incorporated into standard drawing.)

Each child was tested individually in an initial test and a re-test one week later. During both sessions twelve comparisons were presented to the child. Two illustrations at a time were placed side by side on a table and the child pointed to the picture he preferred.

Half of the children were shown one set in the initial test and the other set in the re-test while the other children were presented the sets in reverse order. The order in which the comparisons were presented was also reversed in the re-test. The final control was the placement of the pictures on the opposite side (to the right or left of the child) in the re-test.

Analysis of Results

All the data were analyzed statistically in relation to the controls which had been in effect. The order in which the comparisons were presented did not seem to influence the responses when the preferences from the first half and the last half of the

series were compared.

There were no significant¹ differences between the two sets of pictures when the preferences were compared. There was, however, an interesting finding in relation to the placement of the picture on the right or left. The preference for the picture on the right was significant.

Preferences Right	Preferences Left	N
786 (54.6%)	654 (45.4%)	1440 (100%)
$\chi^2=13.68$	Df=1	P=.01

This bias for the picture on the right was present at the three ages, but there was an observable tendency for the bias to increase with age.

A rank order correlation of the initial results and the re-test results was employed to ascertain the reliability. A surprisingly high correlation was found: $\rho .89$ which becomes $.94$ with the Spearman-Brown correction formula.

Because young children are considered to be quite inconsistent, there was a comparison of all responses made by each child in the two tests. Though many children changed their choices on the re-test, a significantly greater number made the same choice.

Choices the Same	Choices Different	N
393 (55.6%)	327 (44.4%)	720 (100%)
$\chi^2=6.04$	Df=1	P=.02

The stability of preferences did not differ significantly with age. The five year olds, however, tended to make more stable choices than the three's and four's.

To find the preferences for each comparison, all twelve pairs were analyzed individually to see if one illustration was preferred significantly to the other. First, the total responses for each pair in the initial test were tested for significant differences. Likewise the re-test responses and then the combined responses were tested. Secondly, the preferences of the children making the same choice were totaled and this stable group of preferences was analyzed. Finally, the three age groups were compared and tested to see if preferences differed according to age.

Secondary analysis was made with each comparison to see if there were any differences influenced by: sex, socio-economic status, reading habits at home, alertness or activeness of the child, or different nursery school group.

Conclusions

The findings of the study seemed to justify the following conclusions concerning the illustration preferences of the three to five year old children in this investigation:

1. Young children may be inconsistent in their picture preferences but a significant number do make consistent choices.
2. A high reliability is possible with the initial test and re-test technique with young children.
3. There are no significant differences in preferences of the children in this study in relation to: sex, socioeconomic status, research population, reading habits in the home, or alertness or activeness of the child.
4. A significant bias is shown for a picture placed on the right.
5. Light tints in illustrations are significantly preferred to bright saturated colors.
6. Dark shades in illustrations are significantly preferred to bright saturated colors.
7. A photograph is significantly preferred to a black and white line drawing.
8. A fanciful drawing seems to be preferred to a true-to-life drawing though not significantly.
9. A fanciful drawing is significantly preferred to a modified realistic drawing.
10. Three year old children significantly prefer a modified realistic drawing while five year old children prefer a true-to-life drawing when the two are compared.
11. Illustrations with more colors are significantly preferred to those with fewer colors.
12. When a black and white photograph is compared to a line drawing with one color, there does not appear to be any preference for one over the other.

Discussion and Implications for Further Research

The research design used in this study for illustration preferences of young children seems to have merit for future studies of preference. It would seem advisable to use a test-retest technique, not only to establish reliability of the results, but also to allow an analysis of the highly reliable stable responses.

In studies involving the placement of objects to be compared on the right or left, it would be necessary to impose a control so the responses will not be biased for the object on the right. This bias has interesting implications for research concentrating on the observed tendency for the bias to increase with age in three to five year olds. It would be important to see if the bias is consistent with the handedness of the child. Possibly the placement of pictures more often on the right in books has influenced this bias and may be present only in choices of pictures.

The popularity of the lighter and darker colors in preference to the bright colors, being in contradiction to frequently held assumptions and earlier research (1, 3), implies the need for a duplication of this study. A continuation of the study should in-

1. A significant difference has been considered only when the level of confidence has been .01, .02 or .05.

TABLE I
ILLUSTRATION PREFERENCES OF 3, 4, AND 5 YEAR OLD BOYS AND GIRLS.
RESPONSES OF CHILDREN MAKING THE SAME CHOICE ON BOTH
INITIAL AND RE-TEST

Variables	Stable Responses	Significance of Preference*
Light colors	32 (76%)	$x^2 - 10.50$
Bright colors	10 (24%)	Df - 1
Total	42 (100%)	P - .01
Dark colors	21 (75%)	$x^2 - 6.04$
Bright colors	7 (25%)	Df - 1
Total	28 (100%)	P - .02
Photograph	25 (74%)	$x^2 - 6.60$
Black and white drawing	9 (26%)	Df - 1
Total	34 (100%)	P - .01
Photograph	17 (61%)	$x^2 - 0.90$
Drawing with 1 color	11 (39%)	Df - 1
Total	28 (100%)	P - N. S.
Drawing with 1 color	26 (76%)	$x^2 - 8.50$
Black and white drawing	8 (24%)	Df - 1
Total	34 (100%)	P - .01
Drawings with more colors	63 (75%)	$x^2 - 20.00$
Drawings with fewer colors	21 (25%)	Df - 1
Total	84 (100%)	P - .01
Fanciful drawing	22 (69%)	$x^2 - 3.80$
True-to-life drawing	10 (31%)	Df - 1
Total	32 (100%)	P - .05
True-to-life drawing	21 (55%)	$x^2 - 0.24$
Modified realistic drawing**	17 (45%)	Df - 1
Total	38 (100%)	P - N. S.
Fanciful drawing	36 (84%)	$x^2 - 10.40$
Modified realistic drawing	7 (16%)	Df - 1
Total	43 (100%)	P - .01

* Test for the significance of the difference between the two variables.
 x^2 - chi square Df - degrees of freedom P - level of confidence N. S. - not significant

** See Table II

TABLE II
DIFFERENCES IN PREFERENCES FOR MODIFIED REALISTIC
AND TRUE-TO-LIFE DRAWINGS ACCORDING TO AGE
(Stable responses only)

Age	Modified Realistic	True-to-Life	N	
3	9 (90%)	1 (10%)	10	$\chi^2 - 8.70$
4	5 (42%)	7 (58%)	12	Df - 2
5	3 (19%)	13 (81%)	16	P - .02
Total	17	21	38	

χ^2 - chi square Df - degrees of freedom P - level of confidence

clude older children to find if there is the stronger preference for softer colors, as indicated by the seven to nine year olds studied by Martin (4). Bou and Lopez (2) found the preference for the darker shades with their group of Puerto Rican children. The present findings with preference for darker colors by younger American children may imply the desirability of studying the preferences crossculturally. Further studies could compare dark and light colors to each other, as well as to bright colors.

Black and white photographs were not compared to a fully colored line drawing in this study but would be recommended for future investigations. In line with Rudisill's findings (6) the photograph was preferred to the black and white line drawing, and suggests more frequent use in children's books.

The whole area of testing style of drawing should be carefully scrutinized. The definitions of true-to-life, modified realistic and fanciful in this study are over-simplified and questionable. Past studies have used other definitions which may or may not be comparable (2, 3, 5, 6).

Many other interesting aspects of illustrations could be investigated, such as a factor-analysis of illustrations in books that are children's favorites and the effect illustrations have upon children, for instance in dramatic play.

Final Implications

The present analysis, in light of past research, seems to justify these implications:

1. Light and dark colors, as well as the recognized bright colors, may be used effectively in books for young children.
2. More photographs may be used effectively in books for young children.
3. Various styles of drawing should be available

in books for young children. Fanciful elements may be of more interest than currently assumed.

4. More research of children's illustration preferences is needed to confirm the conclusions of this and other studies.
5. Further research of young children's preferences is advisable from a methodological viewpoint, especially in relation to the bias for the right.
6. Preferences should be re-evaluated from time to time.

REFERENCES

1. Bamberger, Florence E., The Effect of the Physical Make-up of a Book upon Children's selection, Johns Hopkins University Studies in Education, No. 4 (Baltimore: Johns Hopkins Press, 1922).
2. Bou, Ismael Roderiguez and Lopez, David Cruz, "Preferences in Colors and Illustrations of Elementary School Children in Puerto Rico," Journal of Educational Psychology, 44, (1953), 490-496.
3. Freeman, G. LaVerne and Freeman, Ruth Sunderlin, The Child and His Picture Book, Chicago: Northwestern University Press, 1933).
4. Martin, Helen, Children's Preferences in Book Illustrations, Western Reserve University Bulletin, New Series 34, No. 10 (Cleveland: Western Reserve University, 1931), 58 pp.
5. Mellinger, Bonnie E., Children's Interests in Pictures, Teachers College Contribution to Education, No. 516 (New York: Teachers College, Columbia University, 1932).
6. Rudisill, Mabel, "Children's Preferences for Color versus Other Qualities in Illustrations," Elementary School Journal, 52, (1951) 444-451.

A SURVEY OF EDUCATIONAL DISABILITY IN EMOTIONALLY DISTURBED CHILDREN*

ARTHUR S. TAMKIN

Columbus (Ohio) Children's Psychiatric Hospital

RESEARCH concerning the processes underlying educational disability in children has demonstrated a complex set of relationships between emotional problems and educational achievement. While most investigators concur in the general finding of a close coincidence of emotional problems and learning disabilities, they often vary with respect to the theories they advance to account for this fact.

Blanchard (1), after evaluating the relations between reading disability and difficulties of personality and emotional development in 73 consecutive cases referred to the Philadelphia Child Guidance Clinic, concluded that there was no causal relationship. She offered the theory that "the reading disability often arises from the same source of difficulty in emotional development, and in much the same manner, as the accompanying personality or behavior problems or neurotic symptoms..."

Support for postulating a causal relationship may be derived from the studies by Fendrick and Bond (3), who found a strong relationship between juvenile delinquency and reading disability, and by Gates and Bond (4), who observed a wide variety of symptomatic reactions attributable to reading disability. The latter evaluated 100 reading disability cases and found that out of this group, approximately three-fourths showed personality maladjustment, of which 25 percent were considered to have personality problems as a result of the disability. Their findings thus permit the inference that, by and large, a learning disability causes emotional and personal maladjustment. It is often pointed out by proponents of this view that a learning disability disrupts the child's functioning in the school situation, making for frustration and difficulties in an important sphere of social adjustment. Severe reading disability, it is maintained, is likely to be a serious catastrophe to many children who feel defeated in a very significant aspect of living. And finally, as Brueckner and Bond (2) pointed out, there are studies which show that some children with severe emotional problems compensate for their difficulties by excessive educational striving, thus becoming extremely competent students.

In addition to the role of emotional and personality problems in educational disability, another important factor is what might be termed an unfortunate home condition. Seigel (6), in his study of personality structure found in reading disability, has called attention to the importance of the home and family relationships to the educational adjustment of the child. Clearly, the significance of the home lies in the fact that family relationships condition the child's ability to follow directions, his use of language, and his background of understanding. Modes of family life which interfere with the orderly development of these basic abilities thereby contribute to educational failings in the school situation.

On the basis of research findings such as these, which demonstrate complex relationships between emotional adjustment, home conditions, and educational disability, one would expect to find rather severe educational disability in children with emotional disturbances requiring residential psychiatric treatment. In this connection Jastak (1946) observed that high reading but low arithmetic scores tend to occur in abnormal mental states of a developmental nature of long standing as in neurosis and schizophrenia in both adults and children.

The purpose of the study reported here was to survey the educational achievement of emotionally disturbed children to ascertain the frequency of educational disability and to determine the relationship between arithmetic and reading skills. It was hypothesized that in a group of hospitalized, emotionally disturbed children, educational disabilities would predominate and that arithmetic would be lower than reading ability.

Method

The Ss consisted of children receiving residential treatment for emotional disorders at a psychiatric hospital. Shortly after admission they were given the Arithmetic and Reading parts of the Wide Range Achievement Test by the teachers of the hospital school.² They were selected for testing on a

* Footnotes will be found at the end of this article.

random basis, and their obtained scores were taken as measures of educational achievement in reading and arithmetic. In all, there were 22 boys and 12 girls with mean ages of 8.7 and 9.4 respectively. The combined mean age was 9.0 years. A special characteristic of this sample and of the population of patients under consideration is that they are devoid of patients with serious medical problems, with psychiatric symptoms requiring isolation and confinement, or with mental deficiency. The absence of mental deficiency was determined either by an IQ measure above 70 or by clinical estimation of the admitting professional staff. Since IQ's were not available for all Ss, chronological age was used as the basis for evaluating educational disabilities in these cases.

Results

The mean grade rating on the Wide Range Achievement Test was 3.48 for the entire sample. For reading alone, it was 3.81 and for arithmetic it was 3.14. The difference between grade ratings for reading and arithmetic is significant at the .005 point based upon a one-tailed test ($t = 2.91$).

In order to evaluate educational disability, the achievement grade ratings were subtracted from the chronological ages of All Ss. Remainders lying between 5.0 and 5.9 would indicate a level of achievement commensurate with the chronological age since a child normally enters the first grade of school between the ages of 5 and 5.9 years. Accordingly, remainders above 5.9 would indicate the presence of disability, while those below 5.0 would indicate advanced achievement. The mean difference between chronological age and grade rating for the sample was found to be 5.44, suggesting that the achievement of these emotionally disturbed youngsters as a whole is commensurate with their chronological age. The mean difference for the boys was 5.28 and 5.73 for the girls. The difference between the sexes is not significant at the .05 level, however.

Observing the difference between chronological age and grade rating for each S, it was noted that 32% of the cases ($n = 11$) demonstrated some degree of educational disability. However, 41% of the cases ($n = 14$) were educationally advanced, and the remaining 27% ($n = 9$) were at their expected grade level.

Discussion

The expectation that educational disabilities would predominate in a group of emotionally disturbed children requiring residential treatment was not borne out by the findings of this study. The Wide Range Achievement Test scores showed that for the group as a whole the academic achievement level was consistent with the expectation for its mean chronological age. Furthermore, an analysis of each individual's achievement level showed that

the bulk of the Ss were either performing at expected grade level or even educationally advanced relative to their chronological ages. Only one-third of the Ss showed signs of some learning disability. Assuming that the Wide Range Achievement Test produces a valid measure of achievement, the findings are most compatible with the theory that educational disability is merely a symptom which might be produced by the same conditions responsible for the emotional disorder. Certainly, emotional problems could not be construed as a cause of educational disability since, as the results demonstrate, only one-third of the emotionally disturbed sample had accompanying educational disabilities.

A possible weakness of this study which could affect the results might be the use of chronological age rather than mental age as the basis from which to measure departures in educational achievement. This choice was necessitated by the unavailability of IQ's for all cases in the sample. The justification for this procedure was predicated upon the assumption that the Ss would not show extreme dispersion about the average IQ range.

The hypothesis that emotionally disturbed children would score higher on the Reading part than on the Arithmetic part was confirmed. Since the mean grade ratings of both parts fell between 3.0 and 4.0, however, it may not be inferred that emotionally disturbed children tend to present arithmetic disabilities since for a group with a mean age of 9.0 years, grade ratings within the obtained range normally would be expected.

Summary and Conclusions

Two hypotheses were presented and tested by means of a survey of the educational achievement of emotionally disturbed children. The hypotheses stated that educational disabilities would predominate and that arithmetic would be lower than reading ability. Thirty-four children recently admitted for residential treatment at a psychiatric hospital were used as Ss, and the Arithmetic and Reading parts of the Wide Range Achievement Test were used to assess educational achievement. Failure to confirm the first hypothesis lent support to the theory that educational disability may be a symptom of the same condition producing the emotional problems rather than being caused by the emotional problems. The second hypothesis was supported by the data, although both arithmetic and reading grade ratings were within the range commensurate with the mean chronological age of the sample.

FOOTNOTES

1. This study was supported by the Ohio Division of Mental Hygiene. The opinions expressed herein are those of the author alone and not necessarily those of the Division of Mental Hygiene.
2. The author is grateful to Miss Jeanne Hugo,

School Principal, who made the Wide Range Achievement Test data available for this study.

REFERENCES

1. Blanchard, P. J. "Reading Disability in Relation to Difficulties of Personality and Emotional Development," Mental Hygiene, New York, 1936, 20, 384-413.
2. Brueckner, L. J. and Bond, G. L. The Diagnosis and Treatment of Learning Difficulties, New York: Appleton-Century-Crofts, 1955.
3. Fendrick, P., and Bond, G. L. "Delinquency and Reading," Pediatrics Seminar and Journal of Genetic Psychology, 1936, 48, 236-243.
4. Gates, A. I. and Bond, G. L. "Failure in Reading and Social Maladjustment," Journal of the National Educational Association, 1936, 25, 205-206.
5. Jastak, J. Wide Range Achievement Test. Manual. Wilmington: Charles L. Story, 1946.
6. Siegel, M. "The Personality Structure of Children with Reading Disabilities as Compared with Children Presenting Other Clinical Problems," Nervous Child, 1954, 10, 409-414.

A STUDY OF PRACTICES AND PROVISIONS FOR THE GIFTED PUPIL IN MATHEMATICS

MONTE SCOTT NORTON
Lincoln, Nebraska, Public Schools

A BASIC PRINCIPLE which underlies the concept of a democracy is recognition of the worth of each individual. In harmony with this principle is the generally accepted principle of optimal educational development on the part of each citizen. Many have seen the need for better educational programs for the gifted. The concern of this paper is with provisions made by Nebraska schools for the gifted in mathematics.

Statement of the Problem

Answers to the following questions were sought: 1) What are some of the actual practices used in Nebraska public high schools to provide for the gifted in mathematics? 2) What are some points of view of administrators in regard to education for the gifted student in mathematics? 3) What are some of the provisions being made in various schools throughout the nation for the gifted pupil in mathematics as shown by a survey of the literature?

Limitations of the Problem

Specifically, this problem dealt with 118 public high schools in the state of Nebraska. All Nebraska public high schools of Class A accreditation and with a minimum total of 125 pupils in their four-year high school program were included in this study along with all Nebraska Class A accredited schools with a three-year high school program; each school in any particular system in Nebraska was considered separately.

The Procedures of this Study

This investigation utilized data from the following sources:

a) Questionnaire. A questionnaire containing pertinent questions concerning practices and provisions for the gifted pupil in mathematics was sent to 118 Nebraska high-school administrators. Part I of the questionnaire consisted of 24 specific questions which the school administrator was asked to answer; each in regard to present practices and provisions for the pupil gifted in mathematics in his

school. Part II of the questionnaire was designed for the purpose of gaining other information on this topic which was not altogether possible through Part I of the questionnaire.

b) Survey of Literature. Through a survey of literature, emerging practices, approaches to problems, opinions, and information concerning programs for the gifted in mathematics were studied. This survey revealed the following facts:

1. Education of the gifted should be the concern of every organized group, institution, and individual in American life. The principle of optimal development for all is in harmony with the democratic way of life.

2. Discovering giftedness in children and youth is the responsibility of parents, teachers, school administrators, and all who live and work with youth.

3. Although the teacher is a vital link in the effectiveness of any gifted-pupil program, the administrator also shares key responsibility in such a program.

4. Enrichment, acceleration, ability grouping, curricular provisions, and extra-curricular activities are the major provisions presently being used to provide for the mathematically gifted student in high schools throughout the nation.

5. Excellent teachers and guidance facilities are paramount in the school's program for gifted pupils.

Of the 118 Nebraska administrators contacted, 104 responded. This means that 88.14 percent of the Class A accredited high schools in Nebraska with a minimum of 125 pupils participated in this study.

The data revealed that Nebraska high schools are making the following provisions for the pupil gifted in mathematics:

1. Acceleration. Responses indicated that 24.3 percent of the Nebraska high schools reporting use acceleration as an administrative provision for the gifted mathematics pupil. That is, 24.3 percent of the schools allow the gifted child to progress more rapidly in work in mathematics than does the average. It is of particular significance to note that many of the administrators mentioned future plans which would include acceleration as a provision for

gifted students.

2. Grouping. Of the 104 schools responding, 55.3 percent use ability grouping in the scheduling of mathematics classes. The majority of schools start making this provision in the ninth grade when pupils are permitted to take algebra 1 as the first step in their advanced mathematics training or to take a course in general mathematics.

3. Curricular provisions. A choice of elective subjects in mathematics designed for the more capable student is provided by 91.3 percent of the schools replying. Advanced algebra, trigonometry, and solid geometry, in that order, completely dominate the mathematical curricular offerings in Nebraska high schools for gifted students. Sixty-one of the administrators stated that gifted pupils in mathematics were permitted to take extra subjects in mathematics in place of free periods. A two or three-track program in mathematics for the full three or four years in high school is offered by 42.7 percent of the schools participating in the study. Many of the smaller high schools reported the use of correspondence work in advanced mathematics when enrollment or personnel limited curricular offerings.

4. Extra-curricular program. One of the weakest areas in regard to the gifted pupil program in mathematics in Nebraska high schools is the extra-curricular program. Only 18.4 percent of the reporting schools are making any extra-curricular provisions for the gifted in their schools. The most frequently mentioned activity in this area was the Mathematics Club. Several administrators reported the use of mathematical contests as a valuable activity for providing for the capable student in mathematics.

5. Preparation of teachers. Thirty-five percent of those administrators reporting stated that they have teachers who are especially prepared for the instruction of the gifted mathematics pupil. The data showed that many Nebraska high school teachers of mathematics have taken special formal

courses in education of the gifted and/or formal courses in advanced mathematics in college or in summer institutes.

6. Library facilities. Library facilities apparently need serious attention. Only 2.9 percent of the administrators rated their school's library as excellent in regard to meeting the needs of the gifted pupil in mathematics. Eighteen of the administrators felt that they had a very good library in this respect. However, 34 percent of the library ratings were fair and/or poor.

7. Teacher load. Limited enrollments and consequent heavy teacher load is a major problem for Nebraska high school educators and citizens as regards the gifted pupil. Frequently administrators checked lack of competent teachers, too heavy a teacher work-load, or too small an overall enrollment as reasons for not offering advanced work in mathematics for gifted pupils. Only 4.9 percent of the Nebraska schools reporting stated that they reduce the class load of teachers who teach pupils gifted in mathematics. Many school officials reporting stated that reorganization plans might solve some of the present problems in this area.

REFERENCES

1. Ahrendt, M. H. "Education of the Mathematically Gifted," Phi Delta Kappan, April, 1953.
2. Baumgartner, Reuben A. "A Mathematics Curriculum for the Gifted," School Science and Mathematics, March, 1953.
3. Educational Policies Commission. Education of the Gifted, Washington, D. C., 1950.
4. Fehr, Howard F. "The Student Gifted in Math," N.E.A. Journal, April, 1954.
5. Passow, Harry. "Talented Youth," Teachers College Record, Columbia University, December, 1955.
6. Price, G. Baley. "A Mathematics Program for the Able," The Mathematics Teacher, October, 1951.

BOOK REVIEWS

L. Joseph Lins, University of Wisconsin; Valworth Plum, University of Minnesota;
John Schmid, Michigan State College

BRAMELD, THEODORE. The Remaking of a Culture (New York: Harper and Brothers, Publishers, 1959), pp. xviii + 478, \$7.50.

This book interprets Puerto Rico's culture in the light of its educational foundations. The method of research is primarily anthropological but includes exploratory techniques. The methodology is worthy of note not only for research relative to the underdeveloped countries but is useful in considering the culture of all countries.

L. J. L.

COMBS, ARTHUR W.; SNYGG, DONALD. Individual Behavior - A Perceptual Approach to Behavior (New York: Harper and Brothers, Publishers, 1959), pp. ix + 521, \$6.00.

This is a revision of the 1949 book which proved very useful in the understanding of human behavior. Both Part I: The Perceptual Frame of Reference and Part II: The Perceptual Approach Applied have new chapters added and former chapters revised and reorganized. The book is recommended for students, teachers, psychologists, and psychiatrists.

L. J. L.

CRONBACH, LEE J. Essentials of Psychological Testing (New York: Harper and Brothers, Publishers, 1960), pp. xxi + 650, \$7.00.

Cronbach has quite completely revised and brought up to date the earlier 1949 text even to the inclusion of the 1960 Stanford-Binet test. The purpose of this textbook is to present the principles of testing in such a way that the student will learn to choose tests according to particular needs and to interpret the results on the basis of the potentialities and limitations of the tests. It is directed to the student in psychology and counseling.

L. J. L.

CYPREANSEN, LUCILE; WILEY, JOHN H.; LAASE, LEROY T. Speech Development, Improvement and Correction (New York: The Ronald Press Company, 1959), pp. vii + 353, \$5.00.

This is a text for beginning courses in methods of speech correction and speech therapy. The student is acquainted with the normal speech development of the child and with speech deviations. Speech testing materials are included as well as methods of speech development through individual and group ac-

tivities. This is an excellent book and is recommended highly.

L. J. L.

DELACATO, CARL H. The Treatment and Prevention of Reading Problems (Springfield, Illinois: Charles C. Thomas, Publisher, 1959), pp. xi + 122.

A neuro-psychological approach is made to the problems of reading. Since it was found that among poor readers the ratio of boys to girls is four to one, an analysis was made of 45 boys who were poor readers. The boys ranged in age from eight to 18 years and were from three states and 26 different schools. The objective was to find how poor readers resemble each other. This is a very interesting work representing an excellent beginning for the early recognition of the potentially poor reader.

L. J. L.

DOWNIE, N. M.; HEATH, R. W. Basic Statistical Methods (New York: Harper and Brothers, Publishers, 1959), pp. xii + 289, \$4.50.

The authors' attempt was to write an introductory text for the social sciences which is brief, clear, involves a minimum of mathematics, and "stresses computation, application, and interpretation." In this, they have been successful. Very few books in the statistics field are as easily read or understood as this book is. The book is divided into nine chapters on descriptive statistics; six on inferential statistics and tests of significance; and three on various correlational techniques, test construction and theory, and nonparametric techniques. Each chapter is followed by a set of problems. There is one section giving formulas, one giving answers to the exercises, and one of reference tables.

L. J. L.

GARRISON, KARL C.; FORCE Jr., DEWEY, G. The Psychology of Exceptional Children (New York: The Ronald Press Company, 1959), vi + 586, \$6.00.

This is the third edition of this introductory textbook which discusses the characteristics, identification, and needs of the exceptional child. It describes the ways in which giftedness and disabilities affect emotional, social, and educational adjustment. The research on which part of the book is based is summarized in a non-technical fashion.

L. J. L.

GARRISON, ROGER H. The Adventure of Learning in College (New York: Harper and Brothers, Publishers, 1959), pp. xii + 270, \$3.25

The author does not claim to have written this work with the intention of its being used as a textbook for a course. It is primarily directed to the college freshman as a sort of person-to-person conversation about learning in college. It has many anecdotes of actual experiences with fictitious names being used. The author emphasizes the differences between secondary education and college experiences pointing out that in college the student is the "chief custodian" of his "own growth and learning".

L. J. L.

GRONLUND, NORMAN E. Sociometry in the Classroom (New York: Harper and Brothers, Publishers, 1959), pp. xviii + 340. \$4.50.

Through sociometric testing, the teacher is able to analyze student's social relations, organize class groups with the purpose of improving pupils' social relations, and evaluate the social development of individual pupils. The book is directed primarily to teachers; however, the material should be of interest to students of psychology, sociology, and social psychology. The introduction presents a description of the sociometric technique. Part I is devoted to the methods of sociometric testing; Part II is directed to the meaning of sociometric results; and Part III is the application of sociometric results to educational problems.

L. J. L.

KAPLAN, LOUIS. Mental Health and Human Relations in Education (New York: Harper and Brothers, Publishers, 1959), pp. xiv + 476, \$5.00.

Mental illness is one of the foremost health problems. The author contends that schools must educate for mental health so the youngsters will learn to work together in wholesome and satisfying ways and will develop the ability to live with themselves and with other persons. The author devotes four chapters to how the schools are concerned with mental health and human relations; four chapters deal with the environmental influences on personality; three chapters are directed to a better understanding of child behavior; and four chapters are on human relations in the classroom.

L. J. L.

Langer, RUDOLPH E., Editor. On Numerical Approximation (Madison, Wisconsin: The University of Wisconsin Press, 1959), pp. x + 462, \$4.50.

This book consists of 21 papers which were

delivered at the Symposium on Numerical Approximation conducted by the Mathematics Research Center, United States Army, at the University of Wisconsin from April 21 to 23, 1958. The discussion of numerical approximation centered around the three general themes of linear approximation, extremal approximation, and algorithms.

L. J. L.

LOREE, M. RAY, Editor. Educational Psychology: Supplementary Text, and Study Questions (New York: The Ronald Press Company, 1959) pp. x + 425, \$4.50.

A total of 32 readings with a full bibliographical footnote are included. These are divided into 14 chapters each initiated with an introductory section and followed by questions on the readings and discussion questions. The chief advantage of this book is the bringing together of the thinking of many persons and the attempt to develop skill in applying psychological knowledge and principles to educational problems.

L. J. L.

MARTIN, WILLIAM E.; STENDLER, CELIA B. Child Behavior and Development (New York: Harcourt, Brace and Company, 1959), pp. xxii + 618, \$6.00.

In the area of child behavior and development, the author draws upon such varied sciences as anthropology, psychology, sociology, biology, pediatrics, and physiology. An attempt to change the pattern of behavior must be based upon an understanding of social, cultural, physical, emotional, and intellectual factors. The book draws upon developmental, comparative, and cross-cultural studies. This is a successor to the 1953 Child Development book.

L. J. L.

MOUSTAKAS, CLARK E. Psychotherapy with Children (New York: Harper and Brothers, Publishers, 1959), pp. xviii + 324, \$5.00.

This book should be useful to students of child development and parent education. There are numerous dialogues presented with explanatory backgrounds and analyses. The development is quite non-technical and encompasses behavior of normal, gifted, handicapped, and disturbed children.

L. J. L.

MULHERN, JAMES. A History of Education: A Social Interpretation (New York: The Ronald Press Company, 1959), pp. vii + 754, \$7.50.

Three types of society - primitive, Oriental, and Western - are discussed in the relationship of economic, social, political, religious, moral, and intellectual factors to social evolution and educational change. This is a second edition, the earlier edition being published in 1946. It is divided into four major parts: (1) Society and Education in the Pre-Renaissance World, (2) Social and Educational Change from the Renaissance to the French Revolution, (3) Social and Educational Change from the French Revolution Until Recent Times, and (4) The Dawn of the New Social and Educational Era.

L. J. L.

OSTROVSKY, EVERETT S. Father to the Child: Case Studies of the Experiences of a Male Teacher with Young Children (New York: G. P. Putnam's Sons, 1959), pp. xviii + 173, \$3.75.

Actual case studies demonstrating the need which children have for the companionship of a father are presented. Ostrovsky traces the consequences of father absence and of variations such as the widow's family, and the 'phantom father.' This book is excellent in bringing the problem to the attention of educators and of showing the need for further research.

L. J. L.

JOURNAL OF EDUCATIONAL RESEARCH

VOLUME 53

MAY, 1960

NUMBER 9

CONTENTS

What Teachers and Prospective Teachers Know About Phonics Generalizations—I. E. Aaron	323
Where Are They Now? A Study of Teacher Supply in Mathematics and Science—Kenneth E. White.....	331
Administration of the F-Scale to a Sample of Elemen- tary School Principals and Teachers—Philip Lambert.....	336
Report of an Evaluation of Curriculum Laboratory Services in a Teachers College—Donald S. MacVean.....	341
A Teaching Prognosis Scale for the Guilford-Zimmer- man Temperament Survey—J. C. Gowan.....	345
Dichotomous Prediction of Student Teaching Excellence Employing Selected CPI Scales—Robert E. Hill, Jr.....	349
Evaluative Attitudes and Behaviors of School Admin- istrators—Carl J. Kleyensteuber.....	352
Do Student Teachers and Supervising Teachers Com- municate with Each Other?—Neal Edmund and Lyle H. Hemink.....	355
Evaluation of an Institute for the Training of Elemen- tary-School Science Resource Teachers— H. Seymour Fowler.....	358

DEMBAR PUBLICATIONS, INC.

Madison 3, Wisconsin

WHAT TEACHERS AND PROSPECTIVE TEACHERS KNOW ABOUT PHONICS GENERALIZATIONS

I. E. AARON
University of Georgia

DESPITE THE widespread agreement that phonics should be taught as a part of the word recognition program in basal reading instruction, questions frequently arise about the extent to which teachers actually do teach these skills. How much teachers know about these skills has bearing upon the extent to which they teach them. Do they know these skills when they enter courses on the teaching of reading, or should the course have as one objective the teaching of the skills themselves as well as how to teach the skills to children?

Purposes of Study

Purposes of the study were:

1. To assess teacher and prospective teacher knowledge of second and third grade difficulty level phonics generalizations at the beginning of a course in the teaching of reading.
2. To determine the relationship between knowledge of phonics generalizations and amount of teaching experience.
3. To determine the relationship between knowledge of phonics principles and educational assignment for experienced teachers or assignment preference for those without teaching experience.

Developing a Test of Phonics Principles Knowledge

A test of phonics principles was devised around eight principles of phonics that ordinarily are taught to children who are working with basal readers on second and third grade difficulty levels. One additional situation was added, that of indeterminate vowels in unaccented syllables. The eight principles are listed below.

1. Vowels in open syllables (those ending in vowels) usually have long sounds.
2. Vowels in closed syllables (those ending in consonants) usually have short sounds.
3. In short words ending in e, the e is silent and the preceding vowel usually is long.
4. When two vowels come together, the first is usually long and the second is usually silent.
5. When c comes before e, i, or y, it usually is soft; otherwise it is hard.

6. When g comes before e, i, or y, it usually is soft; otherwise it is hard.

7. The sounds of vowels followed by r usually are modified.

8. When a is followed by l, the a usually is neither short nor long.

These principles are found in most teacher's guidebooks which accompany basal readers. This particular listing is based upon those given in a widely used book on the teaching of reading (1).

The test consisted of 60 multiple-choice, five-option items. Nonsense words involving the various phonics principles were prepared so that the person taking the test would be forced to put the principles into practice in "recognizing" unknown words. Nonsense words were used because enough real words unknown to college students and experienced teachers could not be found to test knowledge of these principles. An attempt was made to set up a situation for the inexperienced college student and experienced teacher similar to that of the elementary, junior high, or high school pupil who would be using these skills in unlocking new words. An example of one item follows:

"The i in stri has the same sound as: (1) i in light, (2) i in ill, (3) i in charity, (4) i in fir, (5) none of these."

The options on this particular item give a long sound, a short sound, an indeterminate vowel sound, and a sound modified by r. The fifth option was placed on all 60 items in order to allow for silent vowels or some sound other than those listed in the previous four possibilities. For every item asking for the identification of an i sound, the item was identical to the above with the exception of the nonsense word. The same was true for the other vowel letters--a, e, o, u, and y--and the soft and hard c and g letters.

Samples of the types of items are as follows:

1. The a in ___ has the same sound as: (1) a in arm, (2) a in add, (3) a in all, (4) a in gay, (5) none of these.
2. The e in ___ has the same sound as: (1) e in key, (2) e in end, (3) e in maker, (4) e in err, (5) none of these.
3. The i in ___ has the same sound as: (1) i in

light, (2) i in ill, (3) i in charity, (4) i in fir, (5) none of these.

4. The o in ____ has the same sound as: (1) o in throw, (2) o in orb, (3) o in odd, (4) o in work (5) none of these.

5. The u in ____ has the same sound as: (1) u in true, (2) u in cut, (3) u in urn, (4) u in nature, (5) none of these.

6. The y in ____ has the same sound as: (1) y in yellow, (2) y in martyr, (3) y in by (4) y in cylinder, (5) none of these.

7. The c in ____ has the same sound as: (1) c in much, (2) c in cry, (3) c in force, (4) c in cello, (5) none of these.

8. The g in ____ has the same sound as: (1) g in sing, (2) g in mirage, (3) g in rage, (4) g in go, (5) none of these.

All sounds to be identified were contained within accented syllables except for two indeterminate vowels in unaccented syllables. Preliminary instructions to the students explained that nonsense words of more than one syllable were syllabicated and that accented syllables were marked.

The following list presents the 60 nonsense words used in the test. These may be inserted into the eight skeleton items above to reconstruct the entire test.

A Sound	E Sound	I Sound
baze	me'-tul	cra-cli'-ter
gac	a-cler'-nic	ait
ta'-bin	a-lade'	liet
co-fawl'	snec	ta-bin'
ag'-i-mut	a-re'-knew	ag'-i-mat
taul	lea'-tor	sirth
caif	phep'-fiz	flif
keav	niel	stri
lac'-co	a-tron'-o-ber'	kron-gu'-i-ly
cus-ma'to-scop'	r i-wes'-ti	phep-fise'
roab		giz-wer'-nere
daic		bir'-len
		gas-wir'-nere
O Sound	U Sound	Y Sound
moal	dru'-mel	bryc'-til
lac-co'	e-bu'-zit	a-cy'-nic
cas-ma'-tu-scop'	so-lus'	
sto'-lin	thune	
nun-brote'	thun	
woz	mue	
cu-toa'	tas-mo'-to-scup'	
nom'-brate		
C Sound	G Sound	
cas-ma'-to-slop	gise	
gac	cag	
cir	gac	
cy'-tin	geb'-in	

By means of the Hoyt Analysis of Variance Method of Test Reliability (2), a reliability coefficient of .98 was obtained. This was based upon 40 cases

which represented the enrollment in one section of a beginning class in the teaching of reading taught by the writer.

Test Administration

In all cases, the test in mimeographed form was administered to the subjects on the third to fifth class day. This was done to cut down on the possibility of students reading about phonics in textbooks and other sources before taking the test. In most instances, the test was administered on the third day of class. The purpose of the test was explained immediately prior to the administration. Approximately 30 minutes were needed to administer the test to a class, but all persons were permitted to finish.

Subjects Used in the Study

Subjects utilized in this research were all enrollees in the introductory course on the teaching of reading taught on campus at the University of Georgia between March, 1957 and December, 1958. The course was taught eight different times during this period and enrolled 293 persons in all sections. These included 104 teachers with one or more years of experience and 189 teachers-in-preparation with no previous experience. Ninety-seven of the teachers were in grades one through three or planned to teach in those grades, and 120 teachers taught in or planned to teach in grades four through six. The remainder were junior high or high school teachers, administrators, visiting teachers, or special education teachers. Almost all enrollees were residents of Georgia or surrounding states.

Analysis of Test Results

The mean score for the 293 cases in this study was 34.22, and the standard deviation was 8.42. Scores ranged from 13 to 59 items correct of a possible 60 items. Table I presents the distribution of the scores of all subjects.

Exact standards are not available for evaluating the performance of teachers and prospective teachers on the test. The average score of 34.22 items means that the group averaged getting approximately 57 percent of the items correct. The distribution of the scores shows that insofar as performance on this particular test is concerned, the subjects vary considerably in their performance. Of the 293 persons taking the test, only 10 students scored 50 or more. Eighty persons, or 27 percent, scored 40 or more items correct on the test. The remaining 213 persons, or 73 percent, got less than 40 items correct.

Analysis of Performance on Different Types of Items

The items on the test were studied in terms of the particular principles they were prepared to test.

TABLE I
SCORES OF 293 STUDENTS ON PHONICS PRINCIPLES TEST

Score	Frequency	Score	Frequency
10-14	1	35-39	56
15-19	9	40-44	45
20-24	27	45-49	25
25-29	50	50-54	8
30-34	70	55-59	2

TABLE II
PERFORMANCE ON DIFFERENT TYPES OF ITEMS
ON PHONICS PRINCIPLES TEST

Type of Item	Percent Correct
Vowels in open syllables	61
Vowels in closed syllables	49
Short words ending in <u>e</u>	68
Vowel digraphs	61
Soft and hard <u>c</u>	58
Soft and hard <u>g</u>	32
Vowel sound modified by <u>r</u>	68
Vowel <u>a</u> followed by <u>l</u>	85
Indeterminate vowel	51

This was carried out to see if the 293 persons performed better or worse on some particular items or if they performed uniformly on all types of items. Table II summarizes the results of this analysis. It will be noted that only 32 percent of the responses to the items on soft and hard g were correct whereas 85 percent of the responses to the items on vowel a followed by l were correct.

Several possible explanations may be offered to explain why the percentages correct on the various types of items were found. Some of these are that the persons in the study actually know certain principles better than others, that items utilized in the test were more like real words on those items where the subjects scored high than on those items on which the subjects scored low, and that certain principles have fewer exceptions than others. Before complete explanations can be offered, a more thorough investigation of these factors must be undertaken.

Comparing Experience Groupings

In order to determine whether or not experience was associated with performance on the phonics principles test, the 293 cases were divided into four experience groupings: no experience, 1 to 5 years, 6 to 10 years, and more than 10 years. The Welch-Nayer L test was used to test the homogeneity of variances (3). Homogeneity of variances is one of the assumptions underlying the use of the analysis of variance technique, which was used for testing group differences in performance on the phonics principles test.

Table III summarizes the calculations of L_1 . L_1 was found to be larger than the value in Nayer's tables (3), and thus the hypothesis of equal variances was accepted.

Since the variances were found to be homogeneous, the analysis of variance calculations were then in order. Table IV presents the computations of the analysis of variance test of differences in performance among the four experience groupings. The F-value was found to be larger than the tabled value of F at the .01 level. Therefore, the conclusion may be drawn that years of experience are associated with performance on the phonics principles test.

Comparing Educational Assignment or Assignment Preference Groupings

A second comparison was made by means of analysis of variance to determine whether or not expressed preference of grade level of the inexperienced students and grade taught or other educational job held by experienced persons, were associated with performance on the phonics principles test. The 293 cases were divided into primary grades (grades 1 through 3), intermediate grades (grades 4 through 6), junior high and high school grades (grades 7 through 12), and a miscellaneous group which was labeled "other." The latter group included principals, superintendents, instructional

supervisors, speech correctionists, and college teachers. Some teachers reported previous experience at grade levels other than the one in which they presently were assigned. In this comparison, they were classed in accordance with their present assignments.

The homogeneity of variances was again tested by use of the L test. Table V presents the calculations. The L_1 -value of .9935 was found to be larger than the L_1 -value in Nayer's tables, and, therefore, the hypothesis of homogeneity of variances was accepted.

The analysis of variance test was next made to determine whether or not educational assignment or preference was associated with performance on the phonics principles test. Table VI summarizes the computations on this test. The F was not significant at the .05 level. Educational assignment or preference, therefore, was not found to be associated with performance on the phonics principles test.

In interpreting this finding, it should be remembered that more than one half of the 293 enrollees were without experience though they were planning to teach in one of the three grade groups or fill some special educational position represented by the "other" category. For this reason, a third comparison was made, that of examining only those with experience in exactly the same manner as the second comparison involving all 293 cases.

The 104 enrollees who had as much as one year of experience in teaching were next grouped into educational assignment groupings. The four groupings of primary, intermediate, junior high and high school, and "other" were again used. The L test, summarized in Table VII, indicated homogeneous variances.

The F-test showed no significant differences in performance among the educational assignment groupings. Calculations are summarized in Table VIII. As in the case of the total group of 293, the educational assignment of the 104 experienced teachers and educational workers was not found to be associated with performance on the phonics principles test.

Summary and Conclusions

This study attempted to assess teacher and prospective teacher knowledge of phonics generalizations at the beginning of a course in the teaching of reading and to determine the relationship between knowledge of phonics principles and amount of teaching experience and present teaching grade level or preference. Subjects consisted of 293 college students enrolled in courses in the teaching of reading. A five-option, 60-item multiple choice test was constructed to measure knowledge of eight phonics generalizations usually taught at second and third grade reading difficulty levels. The findings were as follows:

- 1) The mean score of correct items on the test was 34.22 of a possible 60, and scores ranged from

TABLE III
 L_1 COMPUTATION FOR TEST OF HOMOGENEITY OF VARIANCES AMONG
 EXPERIENCE GROUPS ON PHONICS PRINCIPLES TEST

Group	N	ΣX	ΣX^2	θ'	Log N	Log θ'	N Log θ'
No experience	189	6,250	217,252	10,572	2.27646	4.02415	760.56435
1-5 Years	42	1,472	55,118	3,528	1.62325	3.54753	148.99626
6-10 Years	24	910	36,656	2,152	1.38021	3.33284	79.98816
Above 10 Years	38	1,395	54,889	3,678	1.57978	3.56561	135.49318
Total	293	10,027	363,915	19,930	2.46687	4.29951	1125.04195

Log $\theta' = 4.29951$ Log $L_1 = \bar{9}.98803$ Log₁ - .9728 Accept
 Harmonic mean - 40.20 k - 4 d. f. - 40

TABLE IV
 ANALYSIS OF VARIANCE OF SCORES IN DIFFERENT EXPERIENCE GROUPS

Variance	D. F.	Sum of Squares	Mean Square	F	Hypothesis
Between	3	843	281.0000	4.07	Reject
Within	289	19,930	68.9619		
Total	292	20,773			

TABLE V

L_1 COMPUTATION FOR TEST OF HOMOGENEITY OF VARIANCES AMONG
EDUCATIONAL ASSIGNMENT AND ASSIGNMENT PREFERENCE GROUPINGS

Group	N	ΣX	ΣX^2	θ'	Log N	Log θ'	N Log θ'
1-3	97	3,371	123,077	5,926	1.98677	3.77276	365.95772
4-6	120	3,955	138,961	8,611	2.07918	3.93505	472.20600
7-12	47	1,708	66,040	3,971	1.67210	3.59890	169.14830
Other	29	993	35,837	1,835	1.46240	3.26364	94.64556
Total	293	10,027	363,915	20,343	2.46687	4.30842	1101.95758

Log $\theta' - 4.30842$

Harmonic mean - 52.45

Log $L_1 - \bar{9}.99716$

k - 4

d.f. - 52

 $L_1 - .9935$

Accept

TABLE VI

ANALYSIS OF VARIANCE OF SCORES IN DIFFERENT GRADE GROUPS

Variance	D. F.	Sum of Squares	Mean Square	F	Hypothesis
Between	3	430	143.3333	2.04	Accept
Within	289	20,343	70.3910		
Total	292	20,773			

TABLE VII

L_1 COMPUTATION FOR TEST OF HOMOGENEITY OF VARIANCES AMONG EDUCATIONAL
ASSIGNMENT OR PREFERENCE GROUPINGS, EXPERIENCED GROUP ONLY

Group	N	ΣX	ΣX^2	θ'	Log N	Log θ'	N Log θ'
1-3	29	1,092	44,052	2,933	1.46240	3.46731	100.55199
4-6	36	1,315	51,643	3,609	1.55630	3.55739	128.06604
7-12	27	967	36,969	2,336	1.43136	3.36847	90.94869
Other	12	403	13,999	465	1.07918	2.66745	32.00940
Total	104	3,777	146,663	9,343	2.01703	3.97049	351.57612

Log $\theta' = 3.97049$ Log $L_1 = 9.98445$ $L_1 = .9648$ Accept
 Harmonic mean - 21.90 $k = 4$ d. f. - 22

TABLE VIII

ANALYSIS OF VARIANCE SCORES IN DIFFERENT GRADE GROUPS,
EXPERIENCED PERSONS ONLY

Variance	D. F.	Sum of Squares	Mean Square	F	Hypothesis
Between	3	150	50.00	-	Accept
Within	100	9,343	934.30		
Total	103	9,493			

13 to 59 items correct. Of the 293 persons taking the test, only 10 students, or 2 percent, scored 50 or more. Only 80 persons, or 27 percent, scored 40 or more items correct. Though no standards of performance have been established, the results would appear to support the contention that a course in the teaching of reading should give attention to teaching the teachers and prospective teachers not only how to teach the phonics generalizations but the principles themselves since very few of the enrollees appear to be well-grounded in these principles.

2) The subjects performed much better on certain of the items than they did on others. Percentages of correct answers to the eight types of items ranged from a low of 32 percent to a high of 85 percent. This appears to have implications for instruction in courses in the teaching of reading. Certain of the phonics generalizations may need more emphasis than others. This area needs further investigation.

3) Years of experience was found to be associated with performance on the phonics principles test. The analysis of variance technique was utilized to compare the 293 cases after they were divided into those with no experience, with from one to five years of experience, with from six to ten years of experience, and those with more than ten years of experience. The F-value was significant at the .01 level. Experienced teachers tended to know more about the principles of phonics than inexperienced teachers. A possible and seemingly logical explanation for this is that teachers with experience had learned something about the principles in teaching them to children.

4) Educational assignment or assignment preference was not associated with performance on the phonics principles test. This was the finding with all 293 subjects and also with the 104 experienced persons. In other words, primary teachers did not

appear to score better on the test than did those at other grade levels, even though primary teachers are frequently the ones who have the main task of teaching these skills to children. Possible explanations for this finding may be that first grade teachers, who were classified among the primary teachers, do very little about teaching these skills to children and might have lowered the overall primary scores. However, inspection of the scores did not support this explanation. Another possible explanation is that teachers at other levels still have to teach these skills because some children in those grades may not have learned them earlier, thus forcing these teachers to learn more about the skills being taught. The fact that approximately 11 per cent of the persons in the study reported some former experience at a level other than the one in which they were classified might have contributed to this finding.

Generalizations from this study should be confined to groups similar to that used in the study--teachers and prospective teachers enrolled in courses in the teaching of reading. The study was undertaken primarily to aid in planning course content in the University of Georgia's introductory course in the teaching of reading. However, since many colleges engaged in teacher preparation also teach such courses, the study appears to have interest on a broader scale.

REFERENCES

1. Betts, E. A. Foundations of Reading Instruction. New York: American Book Company, 1957.
2. Hoyt, C. J. "Note on a Simplified Method of Computing Test Reliability," Educational and Psychological Measurement. 1941, 1, 93-95.
3. Johnson, P. O. Statistical Methods in Research. New York: Prentice Hall, 1949.

WHERE ARE THEY NOW?--A STUDY OF TEACHER SUPPLY IN MATHEMATICS AND SCIENCE

KENNETH E. WHITE

Central Michigan College, Mount Pleasant, Michigan

IN THE SPRING of 1958, the writer conducted a follow-up study of a selected group of graduates of Central Michigan College to obtain information concerning their present status in respect to teaching. The population surveyed included graduates from 1947 through 1956, who had majors or minors in the subject fields of mathematics, chemistry, and physics, and who were at the time of graduation qualified to teach in the public high schools of Michigan.

A questionnaire was sent to each of the 279 subjects included in the survey. Replies were received from 216 and most of the needed information was acquired from campus sources for an additional 14. From these two sources, most of the desired data was acquired for 230 individuals, or 82.4 per cent of the original population.

Before summarizing the results of the study, a few terms need to be defined.

Major: Thirty semester hours of systematic preparation in a field of study.

Minor: Fifteen semester hours of systematic preparation in a field of study.

Currently teaching: Denotes one who was teaching one or more classes at the secondary level (grades 7 through 12) during the spring of 1958.

Results of the Study

Table I summarizes the distribution of the total population included in the survey, by majors, minors, and subject field. Many of those with majors in one of the fields also had minors in one or more of the other fields; however, each person was tabulated only once in the distribution. Those listed under mathematics and physics, mathematics and chemistry, and chemistry and physics had either majors or minors in two fields.

The high point in preparation of teachers in these fields was reached in 1950. Table II shows that 55 were certified for secondary teaching in that year with either a major or a minor or some combination of the two in physics, chemistry, or mathematics. This contrasts sharply with 1956 when only 15 were certified, and 1955 which produced

only 9. It is evident that some very strong factors were at work to produce these inequities since the total number of graduates from the college were about equal for these three years.

The distribution of respondents, in respect to teaching status, is given in Table III. A later analysis indicates that a number of those classed in the non-teaching groups were employed in an educational capacity--high school principals, college instructors, elementary teachers, etc., but they were not, as defined in this study, currently employed in secondary teaching.

Two especially important items are revealed in Table III: 1) The percentage of individuals continuing in secondary teaching decreased rather regularly with increased number of years of service; 2) when the figures in the table are compared with the percent of respondents that have taught at some time or other, 82.4 percent of the total, we find that about one third of the original teaching group was no longer in secondary teaching. If the trend found in Table III is an indication for the profession as a whole, we can expect that less than 50 percent of the beginning teachers in mathematics and science will be teaching seven years later. With such a high rate of attrition, it would appear difficult, if not impossible, to maintain a high level of teaching competence.

Of equal importance with the overall figures of individuals currently teaching, was the distribution of these teachers among the various fields. Table IV provides this analysis.

Only in the case of the respondents with mathematics majors or minors were more than 50 percent currently teaching. It would appear that the factors affecting departure from the teaching field act with considerable force on the teachers in all three areas.

Another facet of the study was aimed at discovering the considerations that influenced those who never taught, and those who left teaching, to make that decision. The reasons for not entering or for leaving teaching are reported in Table V. Several respondents gave more than one reason, but only

JOURNAL OF EDUCATIONAL RESEARCH

TABLE I

NUMBER PREPARED TO TEACH, FROM 1947 THROUGH 1956,
BY SUBJECT FIELD AND MAJOR AND MINOR

Subject Field	Number of Majors	Number of Minors
Chemistry	33	40
Mathematics	119	57
Physics	8	6
Mathematics and Physics	5	3
Mathematics and Chemistry	5	1
Chemistry and Physics		2
Total	170	109

TABLE II

NUMBER PREPARED TO TEACH, BY YEAR, MAJORS, AND MINORS

Year	Number of Majors	Number of Minors	Total
1947	13	6	19
1948	14	10	24
1949	20	13	33
1950	33	22	55
1951	30	17	47
1952	22	13	35
1953	13	4	17
1954	13	12	25
1955	4	5	9
1956	8	7	15
Total	170	109	279

TABLE III
PRESENT STATUS IN RESPECT TO SECONDARY TEACHING

Year	Teaching	Non-Teaching	Percent Teaching
1947	3	13	18.9
1948	8	10	44.4
1949	11	16	40.7
1950	20	25	44.4
1951	16	20	44.4
1952	17	13	56.7
1953	9	5	64.2
1954	14	7	66.7
1955	4	4	50.0
1956	9	6	60.0
Total	111	119	48.3

TABLE IV
PERCENT OF RESPONDENTS CURRENTLY TEACHING, BY FIELD, MAJORS, AND MINORS

Field	Percent of Majors Teaching	Percent of Minors Teaching	Total Percent Teaching
Chemistry	50.0	34.4	41.9
Mathematics	47.4	64.4	52.8
Physics	28.6	75.0	45.4
Mathematics and Physics	40.0	66.7	50.0
Mathematics and Chemistry	20.0	00.0	16.7
Chemistry and Physics	00.0	00.0	00.0
Total	45.8	52.3	48.3

TABLE V
REASONS FOR NOT ENTERING OR FOR LEAVING TEACHING

Reason Reported	Number Reporting	Percent of Non-teaching Group
Financial	32	26.1
Marriage or family	17	14.3
School administration	12	10.1
Other interests	10	8.4
College teaching	7	5.9
Armed forces	7	5.9
Elementary teaching	7	5.9
Lack of desirable placement	5	4.2
Enrolled in graduate school	5	4.2
Deceased	3	2.5
In prison	1	0.8
Teaching load too heavy	1	0.8
Unknown	11	9.2
Total	118	100.0

TABLE VI
PRESENT EMPLOYMENT OF NON-TEACHING GROUP

Occupation	Number	Percent of Non-teaching Group
Employed in Industry (Engineers, researchers, clerical)	24	20.3
Employed in Business (Retail, agents, brokers, etc.)	14	11.9
Professional (Dental, medical, ministry, social work)	8	6.8
Employed by Government	2	1.7
Unemployed	1	0.9

the initial or most important one is included in the analysis.

In Table V, one finds that the three major causes listed accounted for 50.4 percent of the group either not entering or leaving teaching. Of these three, two can be expected to occur in the normal sequence of events. That is, women can be expected to marry, raise families, and consequently leave teaching; we can also expect that a portion of those engaged in teaching will move into administrative positions. The significant figure is the one concerned with the loss of teacher personnel for financial reasons. More than one fourth of the total loss resulted from this factor alone.

It is interesting that such a small percent of the group reported that they left or did not enter teaching because of other interests. More than one half of these are now engaged in some other professional area--namely dentistry, medicine, or the ministry. This low figure would seem to indicate, apart from the women, that the portion of teacher-trainees that look upon the program as an insurance policy is less than many have suspected.

Information was also gathered concerning the length of service of those who have taught but are not currently teaching. It varied from one half year to nine years. The mean number of years was slightly over four. When one recalls that a number of these individuals could not have had this many years of service if they were currently teaching, the figure takes on added significance. Many of those leaving teaching were experienced teachers, and as they indicated, satisfied with teaching itself, but dissatisfied with other considerations, especially the financial rewards!

A review of the present occupations of the non-teaching group is shown in Table VI. Only those employed in non-educational positions are shown.

The tabulation does not include all that were employed. The information could not be obtained for 11. When the reason for not entering or for leaving teaching is compared with current occupations, it is found that almost three fourths of those in business and industry indicated the primary reason that they are not teaching is financial. Oddly, not one of the women who is not teaching has taken a position in some other area.

The final phase of the study dealt with the immediate intentions of the survey group in respect to probable intentions insofar as teaching is concerned. Replies revealed that 8 of those who are currently teaching plan to leave the profession in the near future. However, 6 of these were married women and undoubtedly their return to teaching is dependent upon securing employment within close proximity to their present residence. Thus, it would appear that no significant increase in teacher supply can be expected from the group that has left teaching.

The results of the study appear to establish the following:

- 1) A grave problem exists for the teaching profession in respect to establishing suitable levels of teaching competence because of the high percentage of loss of qualified personnel. Fewer than 50 percent of the respondents, all of whom were fully qualified to teach at the secondary level, are currently engaged in secondary teaching.

- 2) The findings show that the only inducement that would have substantially reduced the loss of personnel would have been improved financial status. The current supply of qualified mathematics, physics, and chemistry teachers would probably be about 25 percent greater if salaries were more suitable.

- 3) Most of the teachers that would have remained in the profession if salaries had been higher are now employed by industry or business. If we are to achieve a high level of education, this drain upon teaching resources cannot be continued.

- 4) Although a few mentioned dissatisfaction with teaching conditions, not one gave this as his primary reason for leaving teaching. One might generalize from this that teachers are as a whole fairly well satisfied with conditions in the schools.

- 5) Although a few that have left teaching plan to return in the near future, the number will scarcely offset the loss of those who have definitely decided to leave teaching. Thus, as far as supply of teachers in the areas of mathematics and science is concerned, barring unusual economic or other conditions, there is little chance that improvement in supply will come from former teachers returning to the classroom.

ADMINISTRATION OF THE F-SCALE TO A SAMPLE OF ELEMENTARY SCHOOL PRINCIPALS AND TEACHERS

PHILIP LAMBERT
University of California, Los Angeles

ELEMENTARY SCHOOL teachers and principals are in a position to have a great effect on the beliefs and values of American youth. It should therefore be both interesting and valuable to discover what basic attitudes are held by this influential group. The present paper attempts to answer the following question: to what extent do elementary teachers and principals have authoritarian attitudes? The first problem was to find some test of authoritarianism; the measure chosen was the F-scale devised by Adorno et al (1). After the test results had been collected and analyzed, the general statistical properties of the F-scale as used for this particular group were studied, and an item analysis of the scale was made.

The F-scale has been administered to a wide variety of groups with the implication that differences in the mean scores reflected differences in underlying potential fascist personality trends.

In 1950 Adorno et al (1) reported a mean score of 4.73, S.D. --.86 for San Quentin prisoners; for working-class men 4.19, S.D. 1.18; for university women students 3.51, S.D. .90; and for women patients at a psychiatric clinic 3.69, S.D. 1.30.

Cohn and Carsch (2) in 1954, testing a group of Germans, found a mean score of 5.26, S.D. .86. They concluded that if one makes the assumption that these German workers are "fascistic" personalities, then the data appear to support the validity of the F-scale.

McGee (3) in 1954 found an overall mean F-scale score of 2.89, S.D. .88 for a group of elementary and secondary teachers. The data supported his hypothesis that the mean F-scale score for a teacher sample would be lower than the average of the 14 groups of middle-class adults reported in *The Authoritarian Personality* by Adorno et al (1), 3.84.

McGee suggested three reasons for this: (a) although the authors of the F-scale tried to limit the scale to statements they believed to be democratic, numerous items are still openly prejudiced; (b) because of their education and training, teachers as a group tend to know the "right" answers; (c) as a group selected from among the population at large, teachers are actually less authoritarian than the average adult. This last explanation seemed to Mc-

Gee to be the most likely one, although he conceded that the other two alternatives probably also contributed to the lower overall mean.

Methodology

Data from elementary school principals were obtained by distributing the F-scale at 13 different principals' meetings. This was done either by the experimenter or by the cooperating superintendents of the various school districts in California. In both cases a letter was circulated at the same time introducing the questionnaire as a public opinion survey. Subjects were guaranteed anonymity. In all, 170 forms were distributed to principals, and 146 (86%) were returned. Of these, 14 (9%) were rejected since 11 forms were unidentifiable and three arrived too late to be included in the study.

Data from teachers were obtained with the cooperation of some of the principals already tested. To 30 of these principals, the experimenter sent F-scale forms and explanatory letters similar to those distributed at the principals' meetings. The 30 principals also received a letter asking for their cooperation in distributing the questionnaires to their teachers and in collecting and forwarding the sealed envelopes which contained the completed forms. In all, 510 forms were distributed to teachers, of which 305 (60%) were returned. Of these, 14 (4.6%) were not scored or used as part of the sample, since seven were unidentifiable, four arrived after the analysis was completed, and three were lost.

The results which follow are based on an analysis of 423 forms, 132 from principals and 291 from teachers.

Results

Data indicating the reliability and related statistical properties of the F-scale are given in Table I. The mean score for all elementary principals and teachers was 3.04, S.D. .97. The split-half reliability of the total scale was .88. This agrees closely with McGee's study on a similar group. He found a split-half reliability of .86 with a mean of

TABLE I
RELIABILITY OF THE F-SCALE SCORES

Group	Number	Reliability ^a	M	MDN.	S. D.	Range
Principals (Male)	98	.85	2.86	2.88	.80	1.27-4.47
Principals (Female)	34	.94	2.97	2.91	1.01	1.33-6.20
Principals	132	.88	2.88	2.89	.86	1.27-6.20
Teachers (Male)	61	.81	2.65	2.71	.67	1.13-4.26
Teachers (Female)	230	.91	3.23	3.25	1.05	1.00-6.00
Teachers	291	.91	3.11	3.06	1.01	1.00-6.00
All	423	.88	3.04	2.99	.97	1.00-6.20

All reliabilities are based on correlations (Pearson r) between the odd items and even items. The correlation coefficients were corrected by the Spearman-Brown formula to give the reliability values in the table.

TABLE II
MEANS AND DISCRIMINATORY POWERS OF THE F-SCALE ITEMS (FORM 30)

Item	Men's Group			Women's Group			Men and Women Combined		
	Mean	Rank	D.P.	Mean	Rank	D.P.	Mean	Rank	D.P.
1. War and conflict	3.71	(7)	3.09	3.68	(7)	2.33	3.70	(7)	2.88
2. Obedience and respect	2.74	(13)	1.65	3.11	(13)	2.56	2.85	(14)	1.91
3. Will power	3.94	(5)	1.83	4.19	(4)	1.33	4.01	(5.5)	1.69
4. Science	4.11	(2)	1.13	5.08	(2)	3.89	4.39	(2)	1.91
5. Supernatural power	2.66	(15)	1.78	3.54	(9)	3.00	2.91	(12)	2.13
6. Cheerful things	2.45	(21)	.52	3.43	(10)	2.56	2.73	(15)	1.09
7. Bad manners	3.98	(4)	2.57	4.11	(6)	1.56	4.02	(4)	2.28
8. Discipline and determination	2.68	(14)	2.52	2.46	(20.5)	2.89	2.62	(16)	2.63
9. Born with urge	2.56	(16)	2.00	2.57	(19)	1.56	2.58	(20)	1.89
10. Infection and disease	2.49	(20)	1.70	2.92	(16)	1.89	2.61	(17)	1.75
11. Honor	2.54	(17)	1.61	2.70	(18)	2.33	2.59	(19)	1.81
12. Rebellious ideas	3.99	(3)	3.13	4.16	(5)	3.45	4.04	(3)	3.22
13. Devoted leaders	2.79	(12)	2.61	3.19	(12)	2.22	2.90	(13)	2.50
14. Sex crimes	2.33	(23)	2.83	2.19	(23)	3.11	2.29	(23)	2.91
15. Weak and strong	1.27	(30)	.61	1.76	(27)	1.33	1.40	(30)	.81
16. Love for parents	2.52	(18)	2.70	2.78	(17)	2.67	2.60	(18)	2.69
17. Astrology	1.74	(29)	.65	2.14	(24)	2.44	1.85	(28)	1.16
18. Prying	3.70	(8)	2.96	3.41	(11)	2.78	3.62	(9)	2.14
19. Earthquake	2.00	(26)	.74	1.70	(29)	1.11	1.92	(26)	.84
20. Immoral people	1.80	(28)	.87	1.92	(25)	1.89	1.83	(29)	1.16
21. Wild sex life	2.16	(25)	2.09	1.84	(26)	1.67	2.07	(25)	1.97
22. Talk less	3.68	(9)	3.39	3.65	(8)	3.11	3.67	(8)	3.31
23. Plots	2.34	(22)	2.39	2.46	(20.5)	2.56	2.37	(22)	2.44
24. Homosexuals	1.91	(27)	1.91	1.73	(28)	1.33	1.86	(27)	1.75
25. No sane person	3.18	(10)	2.78	2.95	(15)	2.33	3.11	(10)	2.66
26. Familiarity	2.91	(11)	1.61	3.05	(14)	2.89	2.95	(11)	1.97
27. Effective work	2.30	(24)	1.87	1.59	(30)	1.00	2.10	(24)	1.63
28. Hard work	4.31	(1)	3.26	5.27	(1)	2.33	4.58	(1)	3.00
29. Seamy side	3.90	(6)	3.30	4.27	(3)	2.56	4.01	(5.5)	2.39
30. Human nature	2.51	(19)	1.83	2.24	(22)	.67	2.44	(21)	1.50
Mean/person/item	2.84	2.06		3.00		2.25	2.89		2.09

TABLE III
MEANS AND DISCRIMINATORY POWERS OF THE F-SCALE ITEMS (FORM 30)

Item	Men's Group			Women's Group			Men and Women Combined		
	Mean	Rank	D.P.	Rank	Mean	D.P.	Rank	Mean	Rank
1. War and conflict	3.50	(5)	3.20	(1)	4.43	(3.5)	3.17	4.23	(3)
2. Obedience and respect	2.60	(16.5)	2.47	(10)	3.66	(11)	3.02	3.44	(11)
3. Will power	3.30	(8)	1.93	(14)	4.43	(3.5)	3.19	4.19	(4)
4. Science	4.27	(2)	2.53	(9)	4.75	(1)	2.53	4.65	(1)
5. Supernatural power	2.73	(13)	2.07	(13)	3.69	(10)	3.00	3.48	(10)
6. Cheerful things	2.61	(15)	.60	(27.5)	3.27	(13)	2.50	3.13	(14)
7. Bad manners	3.19	(9)	1.40	(18.5)	3.89	(8.5)	2.95	3.74	(9)
8. Discipline and Determination	2.48	(20)	2.67	(6)	2.97	(17)	3.21	2.86	(17)
9. Born with urge	1.71	(26.3)	.47	(29)	2.68	(21)	1.66	2.47	(24)
10. Infection and disease	2.60	(16.5)	1.20	(22.5)	2.50	(24)	2.24	2.52	(22)
11. Honor	2.55	(18.5)	1.67	(16)	2.93	(18)	2.79	2.85	(18)
12. Rebellious ideas	3.56	(4)	2.87	(3.5)	4.19	(7)	3.12	4.05	(6)
13. Devoted leaders	2.87	(11)	2.87	(3.5)	3.26	(14)	2.36	3.18	(13)
14. Sex crimes	2.13	(23)	3.07	(2)	3.00	(16)	3.60	2.82	(20)
15. Weak and strong	1.39	(30)	.93	(24)	1.90	(29)	1.66	1.79	(30)
16. Love for parents	2.10	(24)	1.60	(17)	2.59	(23)	2.60	2.48	(23)
17. Astrology	1.65	(29)	.60	(27.5)	2.40	(25)	2.10	2.24	(25)
18. Prying	3.77	(3)	1.27	(20.5)	3.89	(8.5)	2.36	3.87	(8)
19. Earthquake	1.92	(25)	1.27	(20.5)	1.89	(30)	1.10	1.89	(29)
20. Immoral people	1.71	(26.3)	1.40	(18.5)	2.32	(26)	2.74	2.19	(26)
21. Wild sex life	2.16	(22)	.67	(26)	2.11	(28)	1.45	2.12	(28)
22. Talk less	3.32	(7)	2.80	(5)	4.22	(6)	3.91	4.03	(7)
23. Plots	2.55	(18.5)	.33	(30)	3.12	(15)	3.04	3.00	(15)
24. Homosexuals	1.71	(26.3)	1.20	(22.5)	2.24	(27)	2.33	2.13	(27)
25. No sane person	2.92	(10)	2.60	(7.5)	3.40	(12)	3.10	3.30	(12)
26. Familiarity	2.63	(14)	.80	(25)	2.89	(20)	2.53	2.83	(19)
27. Effective work	2.24	(21)	2.40	(11)	2.60	(22)	2.64	2.53	(21)
28. Hard work	4.35	(1)	2.20	(12)	4.70	(2)	3.12	4.63	(2)
29. Seamy side	3.40	(6)	2.60	(7.5)	4.28	(5)	3.91	4.10	(5)
30. Human nature	2.76	(12)	1.87	(15)	2.91	(19)	2.52	2.88	(16)
Mean/person/item	2.69		1.79		3.24		2.68	3.12	

2.89.

Teacher-Principal Differences. The mean for all principals was 2.88, for all teachers 3.11. This agrees with the findings of Adorno et al (p.287) if one makes the assumption that, in general, principals are more highly educated than teachers. Adorno found a low but statistically significant correlation between years of education and E-scale scores, which correlate .77 with F-scale scores.

Male-Female Differences. The difference between the mean scores of male and female principals was only .09. The difference between the means of the male and female teachers, however, was .58. Men teachers and principals as a group were the most constricted in their range of scores (1.13-4.26 and 1.27-4.47) and in their internal variability (S.D. .67 and S.D. .80); that is, the men's answers tended to cluster around a position so that there were few extremely low or high scores. (McGee also found a difference between male and female teachers. Adorno et al, however, found only a slight difference between comparable sex groups). Women teachers and principals were less restricted than the men (1.00-6.00, S.D. 1.05 and 1.33-6.20, S.D. 1.01).

Only a slight difference was found between the means of men and women principals (.09), men teachers and men principals (.21), and women teachers and women principals (.26).

Item Analysis

The function of item analysis and the procedures involved when applied to the F-scale are described in detail by Adorno et al (1).

Data on the item analysis of the principal questionnaires are presented in Table II. The average D.P. of 2.09 for the principals is quite satisfactory. It indicates that on most items there were few low-quarter members who agreed with the statements, few high-quarter members who disagreed more than slightly. It is probable that the present sample did not contain subjects who are actually extreme with respect to the pattern which the F-scale was designed to measure. This circumstance would tend to lower the D.P.'s and the reliability coefficients, and may well be one explanation for the difference between the values reported in this study and those reported in Adorno et al (1). For the 30 items, nine D.P.'s are over 2.5, five are between 2.0 and 2.5 and 11 are between 1.5 and 2.0. Only five D.P.'s are below 1.5, and here the group means are also less than 2.73, so that the D.P. is more significant than it appears. For example, in the combined group, the poorest discriminating item (Item 15) differentiates at better than the one per cent level statistically, $t = 3.53$.

Data on the item analysis of the elementary school teacher questionnaires are presented in Ta-

ble III. The average D.P. of 2.49 is also quite satisfactory. As with the principals, the data indicates that the teacher group does not contain subjects who are extreme with respect to the F-scale.

For the 30 items, 15 D.P.'s are over 2.5, ten are between 2.0 and 2.5, and two are between 1.5 and 2.0. Only three D.P.'s are below 1.5, but here again the group means are less than 2.5.

The observation that D.P.'s are lower, on the average, for men teachers (1.79) than for women teachers (2.68) is consistent with the lower mean for men than for women and the men's narrower range. But, despite the absolute difference in the D.P.'s between men teachers and women teachers, items which work well for one sex tend, in general, to work well for the other. The correlations between the D.P. rank order for the men and that for the women teachers is .79, a value which is high enough to justify our averaging the D.P.'s of the two groups to obtain an overall "order of usefulness" for each item.

Summary and Conclusions

The F-scale was administered to a group of 350 California elementary school principals and teachers. The results showed that:

1. This sample yielded a split-half reliability of .88 and a mean of 3.04.

2. There was a difference between the mean scores of principals and teachers, and between men teachers and women teachers; however, very little difference was found between men principals and women principals.

3. An item analysis of the questionnaire demonstrated that the F-scale discriminates well between the upper and lower quartile of an elementary school principal or teacher population.

In terms of both reliability and internal consistency it may be concluded that the F-scale provides an adequate measuring instrument for studying individual differences in authoritarian-equalitarian personality in teachers and school administrators.

REFERENCES

1. Adorno, T. W., Frankel-Brunswick, E., Levinson, D. S., and Sanford, R. N., *The Authoritarian Personality*. New York: Harper, 1950.
2. Cohn, T. S. and Carsch, H., "Administration of the F-scale to a Sample of Germans," *Journal of Abnormal Social Psychology*, 1954, Vol. 54, p. 471.
3. McGee, Henry Morrison, "Measurement of Authoritarianism and its Relation to Teachers' Classroom Behavior." Department of Education, University of California: *Genetic Psychology Monograph*, 1955, Vol. 52, pp. 89-146.

REPORT OF AN EVALUATION OF CURRICULUM LABORATORY SERVICES IN A TEACHERS COLLEGE

DONALD S. Mac VEAN
Ball State Teachers College, Muncie, Indiana

THERE HAS APPEARED in the last fifteen years a considerable body of literature including a number of theses and dissertations on the subject of curriculum laboratories, but there has been no study which measured the actual use of such a facility nor one which evaluated its services in terms of its primary objectives--assisting the teacher training program.

The curriculum laboratory of the Ball State Teachers College library--upon which this study is based--has been in existence eight years. It was created in response to a need expressed by the faculty for printed curriculum materials; specifically for textbooks, workbooks, teachers' manuals, courses of study, curriculum bulletins, units of work, standardized tests, lesson plans and pamphlets on curriculum which would be useful to the faculty and students involved in the teacher training program. The department has had considerable use from its inception.

The curriculum laboratory is a department of the library and under the general supervision of the chief librarian. Directly in charge is a professionally trained librarian who is also an experienced teacher. In addition, seven student assistants work a total of eighty hours per week doing the routine and clerical work.

Purposes

The purposes of the study were, then, as follows: 1) to determine which individuals or group within the library's total patronage use the curriculum laboratory; 2) to learn which materials are used and for what purposes they are used; 3) to determine if the students and faculty receive the materials and services from the curriculum laboratory which they need to carry on their work; and 4) to determine the contribution of the facility to the primary objective of the college--the training of teachers.

Procedure

A two-week period (including weekends) in February, 1957, was selected to conduct the study. By comparing the attendance record with that of previous years, this was found to be a typical period in this respect, except for class visitations. The latter figure was far below the usual number. It was found also, by comparing the current class schedule with those of the three previous years, that all courses relating to teacher education usually given during the winter quarter were being given at this time.

During the two-week period of the study, a daily record was kept of attendance, of materials used and checked out, and of the number and types of questions asked. In addition, each student who came to the curriculum laboratory was asked to register (but only once during the two-week period).

From the 215 students who registered, thirty were selected by use of the Table of Random Numbers¹ and asked to come to the curriculum laboratory and complete a three-page questionnaire. Twenty-two complied with the request. A questionnaire, slightly different than the one given to students, was sent to all faculty members who had used the curriculum laboratory personally or had sent classes to it during the current school year. Twenty-five questionnaires were sent, twenty-three were returned.

In summary, then, the data reported in this study were derived from: 1) the statistical record of attendance and use; and 2) questionnaires completed by twenty-three faculty members and twenty-two students.

Findings

The statistical record revealed that during the two-week period, 1,562 students and eleven faculty members used the curriculum laboratory in some

1. Found in Tate, Merle, Statistics in Education. New York: The Macmillan Company (1955) pp. 568-69.

way. In the total number of 1,562, there were, however, only 215 different students as shown by the registration list. This fact clearly indicated that the department is used frequently and intensively by a relatively small percentage of the student body during any one school term. One class visited the curriculum laboratory also during the two-week period.

Two hundred and five persons used the curriculum laboratory as a study hall only; i. e., they did not use any of the materials belonging to the department. The curriculum materials in the department were used by the following numbers of students and faculty: textbooks, 289; courses of study, 224; teachers' manuals, 54; workbooks, 31; standardized tests, 5; pamphlets, 12; bibliographies, 4; old textbooks in the historical collection, 3. Other curriculum materials contained in the collection but not used at all during the two-week period are the publications of the Citizenship Education Project and the Sloan Project in Applied Economics, and testing bulletins.

The reference books located in the department and the number of persons that used them are as follows: juvenile encyclopedias, 8; dictionary, 18; list of state adopted textbooks, 7; and textbook publishers' catalogs, 6. Not used were the test publishers' catalogs.

The catalogs available and the number that used each are as follows: author-title catalog for courses of study and textbooks, 6; dictionary catalog for tests, 1; and the subject analytics catalog for units contained in courses of study, 14.

Of the indexes on hand, only Textbooks in Print was used and this only once. In sixteen instances, requests for material had to be referred to other departments of the library.

A total of 255 reference questions were asked during the two-week period. One hundred and fifty-eight of these questions related to the location of materials (many of which would have been unnecessary if the catalogs had been used); sixteen were concerned with the use of the materials; six were about the use of the catalogs; twenty-three were related to sources of materials; and fifty-eight questions were too varied in nature to classify.

The questionnaires furnished both students and faculty an opportunity to state their ideas on the value, use, organization, and administration of the curriculum laboratory. Faculty members felt that the value of the curriculum laboratory rested on the fact that its existence permits students and teachers to become familiar with and to evaluate curriculum publications and to keep up to date with such materials; that the descriptions of school programs found in the courses of study and curriculum bulletins provided teachers with helpful classroom methods and techniques; also that these same publications are an excellent source of material for units and other teaching devices.

The questionnaire contained a list of all the materials in the curriculum laboratory. The faculty

members were asked to check those which they had used. The types of materials, arranged according to frequency of use, are as follows: courses of study, textbooks, teachers' manuals (to accompany textbooks), units of work, workbooks, pamphlets on curriculum, standardized tests, lesson plans, old textbook collection, and the Sloan Project in Applied Economics bulletins.

Even though the faculty had found most of the materials in the curriculum laboratory to be useful, they had numerous suggestions of others which should be included, as follows: 1) books for retarded readers, e.g., the Dolch reading series and other adaptations of classic literature in which the vocabulary is graded; 2) books on curriculum construction and research; 3) a greater variety of courses of study; 4) free and inexpensive pamphlet and illustrative material from businesses and industries; 5) educational games, drill devices, and maps; 6) reprints of articles describing specific teaching techniques; and 7) methods books.

Ten faculty members thought that there were enough new and recent materials in the department; three thought that there were not. Thirteen felt that old editions of textbooks and courses of study should be kept; eleven favored discarding them. Only two persons thought that the materials were organized in such a way that they were inconvenient to use. One felt that a broader classification system should be used for textbooks and courses of study. Four thought that the courses of study should be filed by subject in filing cabinets rather than on open shelves. Less than half of the faculty had used any of the catalogs in the department.

Varying responses were received when faculty members were asked to comment on curriculum laboratory policies and services. They were generally satisfied with the location of the curriculum laboratory but four said that they would use it more if it were nearer the classrooms or the education department. The same number thought that it should be nearer the teaching materials service, a department housing all audio-visual aids other than films. The hours of service were reported as satisfactory except in one instance when it was suggested that the department be open regularly during vacations. (In most cases, the department is open during vacation periods.) Regarding the circulation policy, seven said they would use the material more if they circulated for the same length of time as other library materials. Eleven said that this would make no difference to them. All faculty members indicated that they had received adequate assistance from the professional staff members who work in the department. Nineteen of the twenty who reported also said the same about the assistance given by the student assistants. Two reported that staff assistance given on the week-ends was inadequate.

At present, bibliographies of courses of study, books, and tests are issued by the curriculum laboratory. Suggestions from the faculty of other publications which the department might issue were as

follows: 1) a bibliography of secondary social studies textbooks; 2) reproduction of some of the best work of students in methods classes for more general distribution; 3) lists of free and inexpensive materials; 4) lists of resource units; and 5) a bibliography of materials descriptive of teaching techniques in various subject areas.

Faculty members were asked to state any problems which they had encountered in using the curriculum laboratory. The following were mentioned: 1) interference with work of individual students when classes visit the curriculum laboratory; 2) too much material missing (mentioned twice); 3) lack of material in certain subject areas.

When asked to make suggestions for improvement or extension of services of the curriculum laboratory, the following were listed: 1) broaden the range of materials to include such items as the S. R. A. Life Adjustment Series and other pamphlet series; 2) pick-up and delivery service for the various departments; 3) a greater variety of courses of study; 4) prepare copies of the Indiana state adopted list of textbooks for distribution; 5) setup a classified file of free and inexpensive material for distribution; 6) speed up the ordering and processing of new books and courses of study; 7) provide duplicate copies of some titles; 8) permit free circulation of duplicates; and 9) provide more space for the curriculum laboratory.

The questionnaire completed by students differed only slightly from the one for faculty members. The responses received from students were somewhat different, however.

The students were asked to indicate the purposes for which they had had occasion to use the materials. In order of frequency, the following were listed: to prepare a unit work, to prepare lesson plans, to prepare some other type of teaching assignment, to analyze and compare textbooks, to use the room as a study hall, and to compare courses of study.

When asked to indicate which materials they had had occasion to use, the following were checked: textbooks, courses of study, and teaching units (an equal number of times), lesson plans, workbooks, and teacher's manuals (an equal number of times), resource units, pamphlets and old textbooks (same number), and the juvenile encyclopedias. No student reported that he had used the materials of the Citizenship Education Project or the Sloan Project in Applied Economics. One student suggested that examples of children's written work be kept in the department so that student teachers will know what to expect of pupils. Seven also said that there were not enough new and recent materials in the department. Two thought that the materials were not organized in a way which made them convenient to use.

Seventeen students said they would use the materials more if they circulated for a two-week period as do books in the stacks. All who answered said they had received adequate assistance from the professional staff members and all but one said the same about the student assistants. When asked for

suggestions to improve services, the students recommended the following: 1) duplicate copies of some books; 2) a larger room with more tables and chairs; 3) more shelf labels designating subject groupings; 4) a longer circulation period, especially on weekends; 5) a greater variety of materials in industrial arts and business education; and 6) shelve the textbooks and courses of study together. We were also interested in knowing in connection with which part of the teacher training program the students use the curriculum laboratory most frequently. According to the information on the questionnaires, the students use the curriculum laboratory to complete assignments for the following courses (in order of frequency): participation (Education 342 and 350), student teaching, methods courses, and curriculum courses.

Analysis and Conclusions

1. Since all of the faculty members who completed questionnaires were known to the staff, it was easy to classify them by department and to determine the purposes for which they used the curriculum laboratory or required their students to use it. Included were the following: 1) faculty members in the education department who teach curriculum courses and those who supervise participating students and student teachers; and 2) faculty members in the various academic departments who teach methods courses. As would be expected, and, borne out by the questionnaires, the students using the curriculum laboratory were those enrolled in the courses taught or supervised by the above-mentioned instructors. Specifically, they are students taking the following courses: 1) participation or pre-student teaching (Education 342 and 350), 2) student teaching, 3) methods courses, and 4) curriculum courses. Other known users of the laboratory who were not specifically dealt with in the study are in-service teachers in the region.

2. Most of the materials in the collection were used by both faculty and students at some time or other during the two-week period. This would suggest that there is little "dead wood" on hand as far as any type of material (but not individual volumes) is concerned. As expected, the largest sections were most used, i.e., textbooks, courses of study, units of work, lesson plans, and tests, in that order. The catalogs and printed indexes were little used, except for the subject index to the courses of study. For some reason, both faculty and students expect to locate the resources needed either by direct approach to the shelves and files or with assistance from the staff. This, in turn, explains the relatively large number of directory type questions.

3. The purposes for which the materials were used were easy to discern. In the case of faculty members, it was to familiarize themselves with new curriculum materials; for students, it was to complete various specific assignments.

4. The record of reference questions has an

implication for staffing. Although many of the questions were purely locational in nature and could be answered as well by a trained student as by a librarian, yet a number of questions suggest that a person trained in both curriculum work and library science is needed. Such questions as the following are asked: "Which of the reading series have special helps for the poor reader?" "What are the strengths or weaknesses of the various juvenile encyclopedias?" "Where can I find enrichment materials for my superior pupils?" "How does the Indiana textbook adoption system operate?" "What is the core curriculum?" Such questions are not for student assistants to answer, but should instead have professional attention. The fact that numerous questions have to be referred to other departments would also suggest the advisability of having a librarian on duty part of the time, at least.

5. In light of the comments of both students and faculty, it seems that the range of materials included in the collection is too narrow. The following materials should be added: 1) books for retarded readers including remedial reading textbooks and literature books adapted to the needs of the poor reader; 2) a select group of professional books on curriculum construction and teaching methods; 3) free pamphlet materials in bulk quantities so that faculty and student teachers could have them for classroom use; and 4) duplicates of certain text-

books and courses of study. Immediate attention should also be given the policies on acquisition and processing of materials since three faculty members and seven students complained that there were not enough new and recent materials in the department. There was also some dissatisfaction with the classification expressed.

6. Criticisms of the present location, policies, and services were relatively minor. Together, however, they suggest that the following steps be taken to improve and expand services: 1) that a longer circulation period for both students and faculty be considered; 2) that the department prepare and issue more specialized bibliographies; 3) that the collection be constantly evaluated by the staff to discover and correct deficiencies; 4) that better facilities for the accommodation of classes be provided; and 5) that some solution to the high rate of loss be found.

7. It seems apparent that the curriculum laboratory is a vital part of the teacher training program. The attendance record indicates that it is used frequently and intensively by a relatively small percentage of the total student body during any one school term--those students taking methods courses, participation, and student teaching. Both faculty and students stated that the collection was necessary in order to complete assignments for these courses.

A TEACHING PROGNOSIS SCALE FOR THE GUILFORD-ZIMMERMAN TEMPERAMENT SURVEY

J. C. GOWAN
San Fernando Valley State College
Northridge, California

THE PROGNOSIS of teaching effectiveness through the use of predictors at the time of training is still in the main unsolved despite continuing efforts of a large number of researchers. The use of personality test methods has been a favorite one, as indicated, for example by the work of Ryans (11, 12) in the Teacher Characteristic Study, and by Cook and Callis, (1) in the development of the Minnesota Teacher Attitude Inventory. The search for scales of existing tests which would correlate with teaching effectiveness has also been tried (7, 14). There has been some research in the development of special scales from item analysis of existing tests (3). The use of the Guilford-Zimmerman Temperament Survey in efforts of this kind has also been made (2). An intensive sampling of especially effective teachers located by the Teacher Characteristics Study, with respect to personality factors, which included material from G-Z has also provided background (5). The present paper is an attempt to draw together such material as it relates to the G-Z and to develop a screening scale as a result.

Method

Four parallel item analyses were run on selected groups of teaching candidates at a large university. Since both the characteristics of the population, and the general method of item analysis used have been described elsewhere (6, 10), they will not be repeated here in detail. Items which distinguished superior teachers from candidates in general at the five percent level of confidence, and which were consistent among the groups, were the only ones retained. The result of this procedure was the thirty-item scale given in Table I. Norms for 100 women credential candidates are given in Table II.

The raw correlation between halves of the above scale is .71 which when corrected by the Spearman-Brown formula becomes a reliability of .83. This compares with reliabilities from .75 to .85 reported for various scales of the test in the Manual (8). Validity studies are only tentative at present.

Discussion

One problem in using the Guilford-Zimmerman

as a tool is the absence of validating scales, although recent efforts in this direction are reported (9). Further difficulties in teacher evaluation are that teachers tend to run high on such validating scales as are represented by the "K" scale of the Minnesota Multiphasic or the "Good Impression" scale of the California Psychological Inventory. The significance of this phenomenon and its relation to ego strength has been explored elsewhere (4, 13).

The proposed scale may be a useful addition for identifying teacher candidates and other personnel workers who are to have responsibilities of a professional capacity in dealing with children and the public. The use of this scale for research purposes is invited.

Summary

The development of a teaching prognosis scale on the Guilford-Zimmerman Temperament Survey is described. Items on the scale are identified, and norms and reliability given.

REFERENCES

1. Cook, W. W. et al. Manual of the Minnesota Teacher Attitude Inventory. The Psychological Corporation, New York.
2. Gowan, J. C. and Seagoe, May. "The Guilford Zimmerman and the California Psychological Inventory in Measurement of Teaching Candidates," California Journal of Educational Research, 6: 1: 35 - 38, (January, 1955).
3. Gowan, J. C. and Seagoe, May. "A Teacher Prognosis Scale on the MMPI," Journal of Educational Research, 49: 1: 1-12, (September, 1955).
4. Gowan, J. C. "The Relation of the 'K' Scale of the MMPI to Teaching Personality," California Journal of Educational Research, 6: 5: 208-12, (November, 1955).
5. Gowan, J. C. "A Summary of the Intensive Study of Twenty Highly Selected Elementary Women Teachers," Journal of Experimental Education, 26: 2: 115-124, (December, 1957).
6. Gowan, J. C., and Seagoe, May. "Intercorrelations and Factor Analyses of Tests Given

TABLE I
ITEMS IN THE TEACHING PROGNOSIS SCALE ON THE GUILFORD-ZIMMERMAN

Scale Letter*	Stencil Cut ¹	G-Z Number ²	IBM Number ³
G	-	111	23
	+	121	25
R	-	42	39
	-	87	48
	-	117	54
E	-	60	132
	-	80	136
	+	110	142
O	-	226	166
	-	261	173
F	-	152	181
	-	182	183
	-	167	184
	-	192	189
	-	242	199
	-	267	204
	-	272	205
	-	292	209
	-	154	241
	+	159	242
P	+	199	250
	-	209	252
	+	219	254
	+	224	255
	+	234	257
	+	239	258
	+	264	263
	+	269	264
	+	279	266
	+	295	299
M	+		

* These letters refer to the Guilford-Zimmerman scales. It will be noted that the items are displayed in order of their IBM stencil numbering, that is by scales.

1 A "+" answer means that the more of the high criterion group answered the question "yes". On the IBM stencil "+" will be cut at position 1, and "-" at position 3, with respect to the five possible cuts.

2 This refers to the number of the item on the Guilford-Zimmerman test booklet and answer sheet. As these are numbered across by rows, rather than down by columns, they differ from usual IBM answer sheet order.

3 This refers to the standard IBM machine stencil form 1000A1724, which has five rows of thirty multiple-choice item responses. Numbers over 150 refer to the reverse side of such an answer sheet, which will of course require a second stencil form in which 150 will be subtracted from each number given in this column.

TABLE II
NORMS FOR WOMEN TEACHING CANDIDATES ON THE TEACHING PROGNOSIS SCALE OF THE G-Z

Raw Score	%ile	Raw Score	%ile	Raw Score	%ile	Raw Score	%ile
29	100	23	77	17	31	11	6
28	99	22	69	16	23	10	4
27	95	21	59	15	18	9	2
26	92	20	52	14	16	8	1
25	90	19	48	13	11	7	1
24	83	18	41	12	10	6	1

The mean of this norm group is 19.7 and the standard deviation is 4.86.

- Teaching Candidates," Journal of Experimental Education, 27: 1: 1-22, (September, 1958).
7. Gough, H. S. and Pemberton, W. H. "Personality Characteristics Related to Success in Practice Teaching," Journal of Applied Psychology, 36: 5: 307-9 (October, 1952).
8. Guilford, J. S. and Zimmerman, W. Manual of the Guilford-Zimmerman Temperament Survey. Sheridan Supply Company, Beverly Hills, California.
9. Jacobs, S. and Schlaf, A. "Falsification on the Guilford-Zimmerman Temperament Survey," Psychological Newsletter, 10:3:138-45, (January, 1959).
10. MacLean, M. S. et al. "A Teacher Selection and Counseling Service," Journal of Educational Research, 48:9:669-78, (May, 1955).
11. Ryans, D. G. "The Investigation of Teacher Characteristics," The Educational Record, 1953, 34: 371-96.
12. Ryans, D. G. "Teacher Personnel Research, II: Investigation of Teacher Characteristics," California Journal of Educational Research, 4: 2: 73-82, (March, 1953).
13. Sweetland, A. and Quay, R. "A Note on the 'K' Scale of the MMPI," Journal of Consulting Psychology, 17: 4: 314-16, (August, 1953).
14. Tanner, W. C. Jr. "Personality Bases in Teacher Selection," Phi Delta Kappan, 35: 7: 271-75, (April, 1954).

DICHOTOMOUS PREDICTION OF STUDENT TEACHING EXCELLENCE EMPLOYING SELECTED CPI SCALES

ROBERT E. HILL, JR.
Ball State Teachers College, Muncie, Indiana

THE PROBLEM of predicting teacher competence and the related problem of identifying behaviors and factors likely to contribute to teacher competence are among the most thoroughly investigated matters in educational psychology. Professional journals are replete with the results of research in this area, ideas for further research, methodology, and opinions. Extensive bibliographies relating to these matters are available (2, 4, 5, 6). One of the research directions taken is the identification and quantification of personality characteristics of teachers. It is tacitly assumed that the personality characteristics of the teacher will affect the teacher's behavior, thence student's behavior. It is quite generally known that relatively little correlation appears to exist (for predictive purposes) between teacher competence and such factors as general intelligence and scholastic achievement (3, pp 1-23).

The present study was undertaken to see whether selected psychological traits as measured by the California Psychological Inventory, a relatively new instrument by Harrison G. Gough, could be used effectively in predicting student teacher excellence. The stability of the transition between student teacher excellence and employed or on-the-job excellence is an assumption which is, of course, open to question.

At the request of the director of student teaching, the several faculty student teacher supervisors divided their student teachers at the close of the quarter into two categories which may be thought of as the better and poorer student teachers, or the grade A and B categories. This division formed the criterion dichotomy. It should perhaps be specifically acknowledged that this criterion probably contains the same limitations with respect to validity and reliability which are so widely discussed in the professional literature which deals with this problem area. Three of the CPI scales were used in the analysis. They are Do (dominance), Ac (achievement via conformance), and Py (psychological-mindedness).

The following explanation of these scales and their purposes may be found in the manual which accompanies the CPI. The dominance scale at-

tempts to assess factors of leadership ability, dominance, persistence, and social initiative. The achievement via conformance scale attempts to identify those factors of interest and motivation which facilitate achievement in any setting where conformance is a positive behavior. The psychological-mindedness scale attempts to measure the degree to which the individual is interested in, and responsive to, the inner needs, motives, and experiences of others. It was hypothesized that each of these psychological characteristics would contribute to favorable personality attributes which may be causal factors in the case of the better student teachers.

Method and Results

The subjects used were the 204 student teachers active at Ball State Teachers College during the fall quarter of the academic year 1957-58, and comprised both sexes, including full-time and part-time student teachers. While the group studied is, in a sense, a total population, it is here assumed to have been randomly drawn, which strictly speaking it was not. Each faculty student teacher supervisor administered the CPI to the group of students working under him at his convenience in regularly assigned class periods during the months of October and November, employing the standardized instructions available. Students were assured that individual results would be held confidential, and that the results would have no effect whatsoever upon the grade received in student teaching. Responses were scored using the IBM facilities of the local counseling and testing center.

There appears to be a somewhat general trend (in institutions whose primary responsibility is the preparation of teachers) toward issuing marks in student teaching which might be somewhat consistently higher than the institution's general policy would seem to warrant. More specifically, if the institution operates on an A, B, C, D, F grading plan, a disproportionate number of A's and B's are rather likely in the student teaching area. This situation led the writer to employ the statistical procedure known as discriminatory analysis, which uses an indeterminate number of variables to arrive

at a dichotomous prediction. The procedure may be regarded as a special form of multiple regression. The 5% (.05) level of confidence was used for tests of statistical significance, since it represents a probability which is neither extremely liberal nor extremely conservative with respect to the likelihood of either Type I or Type II errors.

The papers were divided into the A and B groups in accordance with the criterion ratings given by faculty supervisors. Raw score computations led to the following mean scores:

	Do (X_1)	Py (X_2)	Ac (X_3)
A (N = 98)	28.9	11.7	30.3
B (N = 106)	28.3	11.0	28.7

The data were first treated individually for each scale, using the traditional z (CR) approach for large independent samples. The significance of the differences between means was tested by the usual formula. Only X_3 was significant ($z = 2.957$).

Next the discriminatory analysis procedure was employed. This process (as is true of multiple correlation in general) permits maximal weighting of the variables employed, to the end of optimum prediction using all variables.

The resultant function was:

$$Y = -.003744 X_1 + .005057 X_2 + .026643 X_3$$

This function gave a mean for the groups as follows:

$$Y_A = .758257 \quad Y_B = .715202$$

$$Y_{\text{total}} = .736729$$

The Y_{total} may be considered a breaking point. If an individual's score is greater than this (i.e., closer to Y_A than Y_B) he is classified in group A. Similarly, if the individual's score is less than Y_{total} , he is classified in group B.

The data were then treated by a regression analysis of variance:

Regression Analysis of Variance

Sources	ss	df	ms
Regression	2.192466	3	.730822
Residual	48.729103	200	.243646
Total	50.921569	203	

$$F = \frac{.730822}{.243646} = 2.99^*$$

$$df = 3/200$$

* Significant

An analysis of variance of the Y's enables one to estimate the expected percentage of mis-classification.

Analysis of Variance of the Y's

Sources	ss	df	ms
Groups	.094398	3	.031466
Within	2.098068	200	.010490

Total	2.192466	203
-------	----------	-----

$$F = \frac{.031466}{.010490} = 2.99^*$$

$$df = 3/200$$

If we assume the distribution of Y's for the A and B groups normal and equally variable, then the difference between Y means for the groups, when divided by its standard error, yields the t of student's distribution.

$$t = \frac{Y_A - Y_B}{2\sqrt{ms_w} \text{ (of AV of Y's)}} \quad df = N - p - 1$$

Where p = the number of independent variables

Numerically,

$$t = \frac{.043055}{.204842} = .210 \quad df = 200$$

From the tables we may expect an absolute value of t greater than .210 between 80% and 90% of the time. We thus may expect to misclassify individuals between 40 and 45 percent of the time, using the function computed.

Finally, the significance of the weights (b 's) was tested by the formula:

$$t = \frac{b}{\sigma_b} \quad df = N - p - 1$$

Only b_3 was found to be significant.

Subject to the limitations of the study, the following conclusions seem warranted:

1. The discriminant function (although significant at the 5% level) is not sufficiently accurate in its prediction to warrant its use, since mis-classifications may be expected between 40% and 45% of the time.
2. The dominance trait as measured on the CPI is not significant at the 5% level. What contribution the dominance trait does make in this study is negative.
3. The psychological-mindedness trait as measured on the CPI is not significant at the 5% level.
4. The CPI scale which alleges to measure achievement via conformance is significant at the 5% level.

REFERENCES

1. Association for Student Teaching, AST Research

- Bulletin No. 1, State Teachers College, Cedar Falls, Iowa, 1957.
2. Barr, A. S. et al, "Measurement and Prediction of Teacher Efficiency," Review of Educational Research, 25:261-9, June, 1955.
 3. Bennett, C. M., The Relationships Between Responses to Pupil Aggression and Selected Personality Characteristics of Student Teachers, unpublished Doctoral Dissertation, University of Michigan, 1956.
 4. Domas, S. J., and Tiedeman, D. V., "Teacher Competence: An Annotated Bibliography," Journal of Experimental Education, 19:101-218, December, 1950.
 5. Guba, E. G., and Getzels, J. W., "Personality and Teacher Effectiveness: A Problem In Theoretical Research," Journal of Educational Psychology, 46:330-44, October, 1955.
 6. Gough, Harrison G., California Psychological Inventory, Palo Alto, California, Consulting Psychologists Press, 1957.
 7. Ryans, D. G., "The Investigation of Teacher Characteristics," The Educational Record, 34: 370-96, October, 1953.
- References on Multiple Discriminant Analysis
8. Annals of Mathematical Statistics, Volume 18, No. 4, December, 1947.
 9. British Journal of Psychology, Statistical Section, Volume III, Part II, June, 1950.
 10. Harvard Educational Review, Volume 21, No. 2, Spring, 1951.

EVALUATIVE ATTITUDES AND BEHAVIORS OF SCHOOL ADMINISTRATORS

CARL J. KLEYENSTEUBER
University of Wisconsin - Milwaukee

School administrators as a group have a characteristic pattern of values which influences their attitudes and behaviors. Furthermore, groups of school administrators with different backgrounds have patterns of evaluative attitudes and behaviors which differentiate the groups from each other and from corresponding groups of teachers. These conclusions were reached following a recent study of the values held by one hundred twenty-six male administrators and one hundred twenty-six male teachers selected randomly from ninety-six public high schools in Wisconsin.

To serve as a basis of reference, "value" was interpreted in the words of Stagner as a "generalization from a group of closely related attitudes". Thus an "economic value" might be evidenced as a generalization of an educator's attitudes toward efficient business management of school affairs, toward practical or useful subjects in the curriculum, toward other specific attitudes related to utility in everyday situations. A value, then, was an object of reference.

Procedures

Evaluative attitudes were identified through the interview method of investigation by obtaining scores for each individual on the six categories of the Allport, Vernon, and Lindzey, Study of Values (1). The classification is based on Eduard Spranger's Types of Men (2) which holds that each of six "ideal types" of man has a characteristic dominant interest or highest value, namely: (a) theoretical - discovery of truth, (b) economic - utility, (c) aesthetic - form and harmony, (d) social - love of people, (e) political - power, and (f) religious - unity. Further, it holds that no person belongs exclusively to any one type, but that he is a mixture in which there is a dominant value.

Behaviors were determined from scores for each individual in the same six categories on the Practices and Procedures Inventory for Educators, a measuring instrument, expressly developed for the study. Group evaluative attitudes and behaviors were identified for each category by obtaining the mean scores for individuals comprising groups based successively upon differences in backgrounds. Table I represents the mean scores and significance in differences in means on the Study of Values for

all administrators and all teachers respectively, each taken as a large group.

Evaluative Attitudes

The Study of Values is a forced-choice measuring instrument, with an average mean score of 40 for each category. It indicates the relative prominence of the theoretical, economic, aesthetic, social, political, and religious values which are factors in personality. The mean scores in Table I show that the group of administrators are relatively high in the religious, economic, theoretical, and political values, and relatively low in the social and aesthetic values. This would seem to point out that administrators as a group would evaluate practices, functions, objectives, personalities, and situations first as to their worthwhileness from the standpoint of their effect upon the creation of the highest and absolutely satisfying value experience, and last as to their worthwhileness from the standpoint of grace, form, or fitness. In general, this would seem to mean, in terms of the Study of Values, that administrators are more interested in the development of "character traits of high ideals and reverence" than of "unselfishness and sympathy"; that they consider the more important function of education as "its preparation for practical achievement and financial reward" rather than "its preparation for participation in community activities and aiding less fortunate persons"; that they feel it is more important to provide additional laboratory facilities than to promote the study and participation in music and art; and that they deem Henry Ford and Galileo of more interest than Florence Nightingale. The data show that administrators generally favor the first of these alternative choices.

Table I, in addition, indicates that teachers have a pattern of values which is similar in rank order to that of the administrators except for the interchange of the theoretical and economic categories. However, a statistically significant difference in degree exists in the means of one value and a trend exists in another; the group of administrators appears to be more social than the group of teachers and shows a tendency toward valuing less those situations and activities which are theoretical.

Furthermore, the data secured from administrators in a sub-group based on previous science

TABLE I
MEANS BY CATEGORIES OF THE STUDY OF VALUES
FOR ADMINISTRATORS AND TEACHERS

CATEGORY	ADMINISTRATORS	TEACHERS	SIGNIFICANCE
	N = 126 Mean	N = 126 Mean	
Theoretical	41.841	43.103	.20
Economic	42.127	41.706	--
Aesthetic	31.992	32.849	--
Social	38.238	36.317	.01
Political	40.714	41.317	--
Religious	45.087	44.746	--

TABLE II
MEANS BY CATEGORIES OF THE PRACTICES AND PROCEDURES
INVENTORY FOR ADMINISTRATORS AND TEACHERS

CATEGORY	ADMINISTRATORS	TEACHERS	SIGNIFICANCE
	N = 126 Mean	N = 126 Mean	
Theoretical	48.690	49.730	.20
Economic	49.028	50.417	.10
Aesthetic	44.675	43.528	--
Social	55.187	52.091	.01
Political	50.409	52.056	.02
Religious	52.060	52.306	--

teaching experience shows that they are significantly more theoretical in evaluative attitudes and tend to be less economic and aesthetic than (a) administrators as a whole and (b) a sub-group of administrators without science teaching experience. In addition, the same sub-group is significantly less theoretical and more social and religious than a like sub-group of teachers with science teaching experience.

Additional raw data gathered from administrators and teachers in sub-groups based successively upon age, experience, geographical location of school served, enrollment of school served, undergraduate preparation, and military service reveal further significant facts.

Behaviors

The Practices and Procedures Inventory is a forced-choice measuring instrument, with an average mean of 50 for each category. It indicates the behaviors of educators in school and near-school situations relative to the six evaluative attitudes previously discussed. The mean scores in Table II demonstrate that the administrators rank relatively high in social, religious, and political behaviors and relatively low in economic, theoretical, and aesthetic behaviors. This would seem to indicate that, in terms of items on the instrument, administrators select personnel more frequently because of "professional skill and adherence to high principles" than because of "stated philosophy relative to work" and "willingness to accept low salary"; that they usually cover bulletin boards with "pictures of leaders" and "codes of ethics" rather than with "art work;" that they emphasize "respect for others" and "character traits" on reports to parents rather than "efficiency in the use of school time" and "appreciation of excellence"; that they engage more in "research and investigations relative to student and teacher welfare" than in "expenditures" and "marking systems"; and that they spend more time with "exercising authority over students" than in "making education more practical".

Table II indicates further that teachers have a pattern of behaviors relative to Spranger's six value types which has the same relatively high and relatively low categories as that of the administrators.

However, a statistically significant difference in degree exists between the means of two values; the group of administrators appears to be more social and less political in its practices than the group of teachers and there is a tendency for them to be less theoretical and less economic.

Correlations Between Evaluative Attitudes and Behaviors

An examination of the correlations of behaviors with values held demonstrates that administrators as a group tend to follow the attitudes they hold in the economic value better than teachers. For the teachers the highest correlation exists in the aesthetic category. On the other hand, the smallest correlation is found in the social value where administrators and teachers are much more social in practice than in attitude.

Conclusions

1. While similarities among groups are more numerous than differences, groups of administrators with different backgrounds have patterns of attitudes and behaviors which differentiate them as groups from each other and from groups of teachers.
2. The three most frequently appearing factors of differentiation are attitudes relative to the social value and behaviors relative to both the social value and the political value.

REFERENCES

1. Allport, G.W., Vernon, Philip E., and Lindzey, Gardner. Study of Values, Booklet and Manual of Directions (revised edition). Boston: Houghton Mifflin Company, 1951.
2. Spranger, Eduard. Types of Men. Translated from 5th German edition of Lebensformen by Paul J.W. Pigors. Halle: Max Niemeyer Verlag. American agent: Stechert-Hafner, Inc., 31 East 10th Street, New York 3.
3. Stagner, Ross. Psychology of Personality. New York: McGraw-Hill Book Company, Inc., 1948.

DO STUDENT TEACHERS AND SUPERVISING TEACHERS COMMUNICATE WITH EACH OTHER?

NEAL EDMUND

Temple University, Philadelphia, Pennsylvania

LYLE H. HEMINK

State University College of Education, Cortland, New York

IN THIS SURVEY, the authors' purpose was to determine to what extent student teachers in elementary education and classroom supervisory teachers in public schools agree on students' areas of greatest and least success.

Upon completion of a nine-week student teaching experience, 113 students in elementary education were asked to list five areas in which they felt they had experienced greatest success and five areas in which they felt they had experienced least success. Examples of "areas" given by the authors in the questionnaire were: teaching of reading, art, music; discipline; lesson planning; small group activities. Students were not limited to these areas, however.

From these returns, approximately 38% (37) were selected at random. Each of the classroom supervising teachers who had worked with these 37 students was then asked to list the individual student's five areas of greatest success and five areas of least success. All 37 teachers responded.

In no cases was there one hundred percent agreement. The highest percentage of agreement reached concerned student failure rather than student success. However, there was only one such case.

It seems significant, though not altogether unexpected, that success or failure regarding the handling of discipline problems in the classroom was most frequently agreed upon by students and teachers. However, agreement involving student failure was appreciably less than that involving success, in fact, seven in the case of student failure or least success as opposed to twelve in the success category. Problems of discipline are obviously a key concern in elementary school teaching and a stu-

dent's success or failure in dealing with them can be determined, to some extent, readily enough.

While planning lessons and other activities are of equal, if not of superior, importance in the classroom, only ten cases of agreement were reported; five success and five failure. In four cases, students and teachers reached agreement with respect to the students' success in teaching art; in one case, there was student and teacher agreement regarding student failure. In all, cases of agreement totaled 79 out of a possible 370.

The findings reported in Table II indicate that students and teachers agree more frequently on areas of greatest success, though the difference when compared to areas of least success (less than four per cent) seems negligible. Whether considered individually or together, the percentage of agreement must be interpreted as an indication of poor communication. Secondly, and tremendously significant, is the fact that little or no attention was given to the student success or failure in teaching content. Certainly one should expect to see some comment about how poorly or how well a student teaches arithmetic, social studies, or any content area.

Even under the most ideal conditions, a safe assumption would be that people working together could and should communicate more effectively. What one person says is too frequently "heard" another way. Interpretations differ; misunderstandings occur. In this important aspect of teacher preparation, every effort must be made to improve communication and make possible more profitable opportunities for the professional growth of student teachers.

TABLE I
AREAS OF AGREEMENT BETWEEN STUDENT
TEACHERS AND SUPERVISING TEACHERS

Areas of Greatest Success	Number of Cases
Discipline	12
Good Student Teacher/Pupil Relations	10
Written Planning	5
Teaching Art	4
Teaching Reading	3
Poise	3
Meeting Individual Children's Needs	3
Understanding Children	2
Total	42
Areas of Least Success	Number of Cases
Discipline	7
Lesson Summaries	5
Written Planning	5
Chalkboard Writing	5
Effective Voice	5
Awareness of Time in Teaching	4
Teaching Music	3
English Usage	2
Teaching Art	2
Total	1
	37

TABLE II
 AGREEMENT BETWEEN STUDENT TEACHERS AND SUPERVISING TEACHERS

Areas of Greatest Success	Number of Cases	Percentage of Agreement
No agreement between student and teacher	5	0
Agreement between student and teacher in one area	13	20
Agreement between student and teacher in two areas	11	40
Agreement between student and teacher in three areas	8	60
Mean percentage of agreement		31.89
Areas of Least Success		
No agreement between student and teacher	3	0
Agreement between student and teacher in one area	21	20
Agreement between student and teacher in two areas	9	40
Agreement between student and teacher in three areas	3	60
Agreement between student and teacher in four areas	1	80
Mean percentage of agreement		21.11

EVALUATION OF AN INSTITUTE FOR THE TRAINING OF ELEMENTARY-SCHOOL SCIENCE RESOURCE TEACHERS

H. SEYMOUR FOWLER
Pennsylvania State University
University Park, Pennsylvania

During the Summer of 1959, a six-week Science Institute, supported by the National Science Foundation, was sponsored by the Pennsylvania State University. Forty-five participants were invited to attend the Institute and were given generous financial support under the provisions as described by the National Science Foundation. Over 900 requests for information concerning the institute were received. In the period preceding the Institute, 450 teachers, administrators, and supervisors completed application and enrollment forms for the Institute. A Selections Committee composed of members of the faculties of the various Departments of the Sciences and Education assumed responsibility for choosing the 45 participants. The Elementary-School Teachers, Supervisors, or Administrators were chosen from those persons who met the following criteria:

1. The participant must be of such an age that it will be possible for him to be active in teaching for a number of years. Hence, the participant, will be able to contribute to a considerable degree to the development of an effective program in Elementary-School science.
2. The participant must have completed at least three years of successful teaching experience and must be employed as a teacher, supervisor, or school administrator during the previous school year.
3. The participant must be recommended by his Superintendent or Principal.
4. The participant must accept the responsibilities of a Science Resource Teacher in his home school upon the completion of the program.
5. The school administrator from the home school of the participant must agree to allow the participant to act as a Science Resource Teacher.

Of course, other subjective criteria for selection were also employed. It might be of interest to the reader to know the nature of the group of participants. The institute group was made up of 35 men and 10 women. Their ages ranged from 24-49 years with a mean age of 33. The National Science Foundation gives dependency allowances for a number of

dependents up to 4. Based on this limitation our group had a mean number of dependents of 2.6. Of the 45 participants, 16 were Elementary-School Principals or Teaching Principals, 24 were classroom teachers, 3 were Supervisors of Curriculum for Elementary School Districts, and 2 were science teachers in Elementary-School systems. The participants had taught from 3 to 28 years with a mean of 6.9 years of experience. The participants were asked to list the number of professional societies, either educational or scientific, in which they held membership. The range in this case was from 0 to 11 with a mean of 3.3. Participants were asked to list the names of professional journals which they read regularly. In this case the range was from 1 to 7 with a mean of 3.9. The participants were asked to list the number of college credits completed in 5 different areas of science. The table below summarizes the participants background:

Semester Hours of College Credit Completed

Subject	Range	Mean
Biological Science	0-43	12.6
Chemistry	0-28	4.8
Physics	0-24	3.2
Mathematics	0-30	5.8
Earth Science	0-17	3.5
Total Science	3-87	29.9

Prior to the Institute it was decided that our program would be planned to prepare persons to accept the responsibilities of a Science Resource Teacher in their home schools. This seemed reasonable since teachers especially trained in Physical Education, Art, and Music are common in Elementary-schools. The Institute planners, then, proposed a

program which would prepare persons to act as Science Resource Teachers. These persons would be expected to:

1. Provide leadership in science curriculum planning.
2. Conduct in-service training programs in science for teachers in their own building or in the school system.
3. Give aid to individual teachers of their system in the form of subject matter, help in setting up real experiments at the elementary level and provide other forms of help in solving the problems in the development of a sound science program.
4. Identify school needs in terms of equipment and reference materials and provide leadership in obtaining, organizing and maintaining these materials.
5. Maintain contact with professional organizations concerned with improvement of science education at the elementary level.
6. Develop, with their colleagues, instructional materials in science for the local situation.

Instructors for the Institute program were competent Scientists and Science Educators. Participants received concentrated training in biological sciences, earth science, physical science and methodology. We of the staff were interested in learning of the participants growth in achievement in general science. We decided to use the Read General Science Test published by the World Book Company as an instrument to measure achievement. This test was constructed to measure the extent to which students have achieved the important objectives of a high school course in general science. The test has two comparable forms AM and BM, each comprising 75 test items carefully selected on the basis of curricular validity and satisfaction of statistical requirements.

Form AM of the Read General Science Test was administered to the group of participants on the evening prior to beginning of instruction. Six weeks later on the evening of the last day of instruction Form BM was administered. The following are data from these tests:

	Form	
	AM	BM
Range of Standard Scores	102-141	107-146
Mean Standard Score	124	134
Standard Deviation	9.1	11.7
Standard Error of Mean	1.35	1.74
Number of Participants	45	45

We employed the "t" test on the results of the two examinations. The "t" determined as a comparison of correlated samples was 9.9. This would indicate a very significant gain in the achievement in general science of the participants when measured by these tests which were administered prior to and after participation in the Institute program.

The Elementary-school teacher who acts as a science-resource-teacher must have a good working knowledge of general science. This development of a working knowledge of science had been proposed as one of the objectives of our Institute. If the results of our achievement tests are an indication of this growth in science competence, then our Institute program achieved one of its objectives.

INDEX

Listed Under Articles and Authors

- Aaron, I. E., The Relationship of Selected Measures to Spelling Achievement at the Fourth and Eighth Grade Levels, December, 1959, 138-143.
- Aaron, I. E., What Teachers and Prospective Teachers Know About Phonics Generalizations, May, 1960, 323-330.
- A Comparative Analysis of Personality and Ego Strength Test Scores for In-Prison, Neuro-Psychiatric and Typical Individuals, Russell N. Cassel, October, 1959, 43-53.
- A Comparative Study of Achieving and Underachieving High School Boys of High Intellectual Ability, Edward Frankel, January, 1960, 172-180.
- A Comparison of Forced vs Free Q-Sort Procedure, Robert D. Hess et al., November, 1959, 83-90.
- A Comparison of Normal and Emotionally Ill Children on the Kahn Test of Symbol Arrangement, Howard H. Fink et al., September, 1959, 35-36.
- A Comparison of the Problems of Certain Anglo- and Latin-American Junior High School Students, Paul Witherspoon, April, 1960, 295-299.
- Acceptance and Rejection as Related to Length of School Attendance, Jack R. Frymier, November, 1959, 112-114.
- Administration of the F-Scale to a Sample of Elementary School Principals and Teachers, Philip Lambert, May, 1960, 336-340.
- Amsden, Ruth Helen, Children's Preferences in Story Book Variables, April, 1960, 309-312.
- Analysis of Observed Traits of Teachers Rated Superior in Demonstrating Creativeness in Teaching, Jesse A. Bond, September, 1959, 7-12.
- An Analysis of Test Scores and Grades for Predicting Success of College Students in English Composition, William E. Kunhart et al., October, 1959, 79.
- An Empirical Study of Performance in Mathematics and Performance on Selected Entrance Examinations, Joseph R. Marches, January, 1960, 181-187.
- An Investigation of an Experimental First Grade Phonics Program, Richard H. Bloomer, January, 1960, 188-193.
- An Investigation of Six Correlates of Student Withdrawal from High School, Joseph C. Bledsoe, September, 1959, 3-6.
- An Investigation of the Effects of a Seventh and Eighth Grade Core Program, Bernard Schwartz, December, 1959, 149-152.
- An Opinionaire on Why College Students Choose to Teach, J. Marc Jantzen, September, 1959, 13-18.
- Are Present Reading Tests Valid for Both Girls and Boys? Walter J. Pauk, March, 1960, 279-280.
- A Research in Mental Arithmetic Involving Subtraction, Herbert T. Olander et al., November, 1959, 97-102.
- A Study of Practices and Provisions for the Gifted Pupil in Mathematics, Monte Scott Norton, April, 1960, 316-317.
- A Study of Predictive Validity of the Minnesota Teacher Attitude Inventory, Harry P. Day, September, 1959, 37-38.
- A Study of Student Disciplinary Practices in Two Georgia High Schools, Karl C. Garrison, December, 1959, 153-156.
- A Study of the Effectiveness of a Summer Remedial Course in English for College Freshmen, Martin L. Zeigler et al., October, 1959, 76-78.
- A Study of Twenty Slow Learners, Sister Agnes Jerome, September, 1959, 23-27.
- A Survey of Educational Disability in Emotionally Disturbed Children, Arthur S. Tamkin, April, 1960, 313-315.
- A Teaching Prognosis Scale for the Guilford-Zimmerman Temperament Study, J. C. Gowan, May, 1960, 345-348.
- Authoritarian Trends in Personality as Related to Attitudinal and Behavioral Traits of Student Teachers, Joseph A. Del Popolo, March, 1960, 252-257.
- Beck, Harry S., The Relationship of Colors to Various Concepts, January, 1960, 194-196.
- Bendig, A. W., Extraversion, Neuroticism, and Student Achievement in Introductory Psychology, March, 1960, 263-267.
- Bledsoe, Joseph C., An Investigation of Six Correlates of Student Withdrawal from High School, September, 1959, 3-6.
- Bloomer, Richard H., An Investigation of an Experimental First Grade Phonics Program, January, 1960, 188-193.
- Bloomer, Richard H. et al., The Effect of Sonic Emotion on the Immediate Recall of Nonsense Syllables, March, 1960, 268-272.
- Bond, Jesse A., Analysis of Observed Traits of Teachers Rated Superior in Demonstrating Creativeness in Teaching, September, 1959, 7-12.
- Cassel, Russel N., A Comparative Analysis of Personality and Ego Strength Test Scores for In-Prison, Neuro-Psychiatric and Typical Individuals, October, 1959, 43-52.
- Changes in Affect Attributable to Instruction in Reading Improvement at the College Level, Meryl E. Englander, February, 1960, 231-236.
- Changes in Ethnic Reaction Tendencies During High School, John D. McNeil, January, 1960, 199-200.
- Children's Preferences in Picture Story Book Variables, Ruth Helen Amsden, April, 1960, 309-312.
- College Students' Ability to Evaluate Their Performance on Objective Tests, Earl W. Kooker et al., October, 1959, 69-72.

- Costin, Frank, Measuring Attitudinal Outcomes of Child Psychology with the Parental Attitude Research Instrument, April, 1960, 289-294.
- Cummins, Robert E., Research Insights into the Relationship Between Teachers' Acceptance Attitudes, Their Role Concepts, and Students' Acceptance Attitudes, January, 1960, 197-198.
- Day, Harry P., A Study of Predictive Validity of the Minnesota Teacher Attitude Inventory, September, 1959, 37-38.
- Del Popolo, Joseph A., Authoritarian Trends in Personality as Related to Attitudinal and Behavioral Traits of Student Teachers, March, 1960, 252-257.
- Dichotomous Prediction of Student Teaching Excellence Employing Selected CPI Scales, Robert E. Hill, Jr., May, 1960, 349-351.
- Do Student Teachers and Supervising Teachers Communicate with Each Other? Neal Edmund et al., May, 1960, 355-357.
- Edmund, Neal et al., Do Student Teachers and Supervising Teachers Communicate with Each Other?, May, 1960, 355-357.
- Effect of Familiar Background Music Upon Film Learning, John Freeman et al., November, 1959, 91-96.
- Effect of Sonic Emotion on the Immediate Recall of Nonsense Syllables, Richard H. Bloomer et al., March, 1960, 268-272.
- Efficacy of Two Tests in Differentiating Potentially Low from Average and High First Grade Achievers, Elmer F. Morgan, April, 1960, 300-304.
- Englander, Meryl E., Changes in Affect Attributable to Instruction in Reading Improvement at the College Level, February, 1960, 231-237.
- Entwisle, Doris R., Evaluations of Study-Skills Courses: A Review, March, 1960, 243-251.
- Evaluation of an Institute for the Training of Elementary-School Science Resource Teachers, H. Seymour Fowler, May, 1960, 358-359.
- Evaluations of Study-Skills Courses: A Review, Doris R. Entwisle, March, 1960, 243-251.
- Evaluative Attitudes and Behaviors of School Administrators, Carl J. Kleyensteuber, May, 1960, 352-354.
- Extraversion, Neuroticism, and Student Achievement in Introductory Psychology, A. W. Bendig, March, 1960, 263-267.
- Factors Affecting the Development of Cooperation in Children, Alex F. Perrodin, April, 1960, 283-288.
- Field Trip as a Supplement to Teaching: An Experimental Study, Joseph S. Hillson et al., September, 1959, 19-22.
- Fifty Seventh Graders: A Comparison of Their Reading Achievement and Expected Achievement in Grades One Through Seven, Jeanne S. Hoyt et al., January, 1960, 163-171.
- Fink, Howard H. et al., A Comparison of Normal and Emotionally Ill Children on the Kahn Test of Symbol Arrangement, September, 1959, 35-36.
- Fitting Asymmetric Student Grade Distributions, Robert H. Riffenburgh, December, 1959, 123-129.
- Foss, Maurice F., Placement Bureau Data and Teaching Success, March, 1960, 276-278.
- Fowler, H. Seymour, Evaluation of an Institute for the Training of Elementary-School Science Resource Teachers, May, 1960, 358-359.
- Frankel, Edward, A Comparative Study of Achieving and Underachieving High School Boys of High Intellectual Ability, January, 1960, 172-180.
- Freeman, John et al., Effect of Familiar Background Music Upon Film Learning, November, 1959, 91-96.
- Frymier, Jack R., Acceptance and Rejection as Related to Length of School Attendance, November, 1959, 112-114.
- Frymier, Jack R., The Relationship of Authoritarianism to Rejection, September, 1959, 33-34.
- Garrison, Karl C., A Study of Student Disciplinary Practices in Two Georgia High Schools, December, 1959, 153-156.
- Gowan, J. C., A Teaching Prognosis Scale for the Guilford-Zimmerman Temperament Study, May, 1960, 345-348.
- Gray, William S., Summary of Investigations Relating to Reading July 1, 1958 to June 30, 1959, February, 1960, 203-222.
- Halfter, Irma T. et al., Measurement of College Level Reading Competence in a Content Area, February, 1960, 223-230.
- Hess, Robert D. et al., A Comparison of Forced vs. Free Q-Sort Procedure, November, 1959, 83-90.
- Hill, Robert E., Jr., Dichotomous Prediction of Student Teaching Excellence Employing Selected CPI Scales, May, 1960, 349-351.
- Hillson, Joseph S. et al., The Field Trip as a Supplement to Teaching: An Experimental Study, September, 1959, 19-22.
- Hoyt, Jeanne S. et al., Fifty Seventh Graders: A Comparison of Their Reading Achievement and Expected Achievement in Grades One Through Seven, January, 1960, 163-171.

Intellectual and Personality Characteristics of University of Utah Students, Frank B. Jex et al., November, 1959, 118-120.

Jantzen, J. Marc, An Opinionaire on Why College Students Choose to Teach, September, 1959, 13-18.

Jerome, Sister Agnes, A Study of Twenty Slow Learners, September, 1959, 23-27.

Jex, Frank B. et al., Intellectual and Personality Characteristics of University of Utah Students November, 1959, 118-120.

Jones, Reginald Lanier, Publication Activities of Members of the Division of School Psychology of the American Psychological Association, April, 1960, 305-308.

Josephine, Sister, C. J. S., Survey of the Research Related to the Reading Ability of the Gifted, February, 1960, 237-239.

Karam, Irvin Albert, Merit-Rating Salary Plans in Public School Systems of the United States, 1955-1956, December, 1959, 144-148.

Kleyensteuber, Carl J., Evaluative Attitudes and Behaviors of School Administrators, May, 1960, 352-354.

Kooker, Earl W. et al., College Students' Ability to Evaluate Their Performance on Objective Tests, October, 1959, 69-72.

Kunhart, William E. et al., An Analysis of Test Scores and Grades for Predicting Success of College Students in English Composition, October, 1959, 79.

Lambert, Philip, Administration of the F-Scale to a Sample of Elementary School Principals and Teachers, May, 1960, 336-340.

Lehmann, Irvin H., Rural-Urban Differences in Intelligence, October, 1959, 62-68.

Lohmann, Kaj et al., Some Perceptual Changes During Sensitivity Training, September, 1959, 28-32.

MacVean, Donald S., Report of an Evaluation of Curriculum Laboratory Services in a Teachers College, May, 1960, 341-344.

Marches, Joseph R., An Empirical Study of Performance in Mathematics and Performance on Selected Entrance Examinations, January, 1960, 181-187.

McCall, William A. et al., Measurement of Teacher Merit for Salary Purposes, October, 1959, 73-75.

McNeil, John D., Changes in Ethnic Reaction Tendencies During High School, January, 1960, 199-200.

Measurement of College Level Reading Competence in a Content Area, Irma T. Halfter et al., February, 1960, 223-230.

Measurement of Teacher Merit for Salary Purposes, William A. McCall et al., October, 1959, 73-75.

Measuring Attitudinal Outcomes of Child Psychology with the Parental Attitude Research Instrument, Frank Costin, April, 1960, 289-294.

Merit-Rating Salary Plans in Public School Systems of the United States, 1955-56, Irvin Albert Karam, December, 1959, 144-148.

Morgan, Elmer F., Efficacy of Two Tests in Differentiating Potentially Low from Average and High First Grade Achievers, April, 1960, 300-304.

Norton, Monte Scott, A Study of Practices and Provisions for the Gifted Pupil in Mathematics, April, 1960, 316-317.

Olander, Herbert T. et al., A Research in Mental Arithmetic Involving Subtraction, November, 1959, 97-102.

Palmer, Charlene D., Principles of Child Growth and Development, March, 1960, 273-275.

Pauk, Walter J., Are Present Reading Tests Valid for Both Girls and Boys?, March, 1960, 279-280.

Perrodin, Alex F., Factors Affecting Development of Cooperation in Children, April, 1960, 283-288.

Placement Bureau Data and Teaching Success, Maurice F. Foss, March, 1960, 276-278.

Prediction of Success in Graduate School at Rutgers University, Milton M. Schwartz et al., November, 1959, 109-111.

Principles of Child Growth and Development, Charlene D. Palmer, March, 1960, 273-275.

Publication Activities of Members of the Division of School Psychology of the American Psychological Association, Reginald Lanier Jones, April, 1960, 305-308.

Reading Skill as a Predictor of College Achievement, Malcom H. Robertson, March, 1960, 258-262.

Relationship of Authoritarianism to Rejection, Jack R. Frymier, September, 1959, 33-34.

Relationship of Colors to Various Concepts, Harry

S. Beck, January, 1960, 194-196.
 Relationship of Selected Measures to Spelling Achievement at the Fourth and Eighth Grade Levels, I. E. Aaron, December, 1959, 138-143.
 Report of an Evaluation of Curriculum Laboratory Services in a Teachers College, Donald S. MacVean, May, 1960, 341-344.
 Research Insights into the Relationship Between Teachers' Acceptance Attitudes, Their Role Concepts, and Students' Acceptance Attitudes, Robert E. Cummins, January, 1960, 197-198.
 Riffenburgh, Robert H., Fitting Asymmetric Student Grade Distributions, December, 1959, 123-129.
 Robertson, Malcolm H., Reading Skill as a Predictor of College Achievement, March, 1960, 258-262.
 Rural-Urban Differences in Intelligence, Irvin J. Lehmann, October, 1959, 62-28.
 Schwartz, Bernard, An Investigation of the Effects of a Seventh and Eighth Grade Core Program, December, 1959, 149-152.
 Schwartz, Milton M. et al., Prediction of Success in Graduate School at Rutgers University, November, 1959, 109-111.
 Sex Differences Among High-School Seniors, Pairlee J. Stinson et al., November, 1959, 103-108.
 Socioeconomic Status, Urbanism and Academic Performance in College, Norman F. Washburne, December, 1959, 130-137.
 Some Attitudinal Differences Among Educational Specialists, Administrators and Teachers, Bartholomew D. Wall, November, 1959, 115-117.
 Some Perceptual Changes During Sensitivity Training, Kaj Lohmann et al., September, 1959, 28-32.
 Some Results of an Enrichment Program for Gifted Ninth Graders, J. A. R. Wilson, December, 1959, 157-160.
 Summary of Investigations Relating to Reading July 1, 1958 to June 30, 1959, William S. Gray, February, 1960, 203-222.
 Survey of the Research Related to the Reading Ability

ity of the Gifted, Sister Josephina, C. J. S., February, 1960, 237-239.
 Stimulating Interest in Public Health Problems Among High School Pupils, Helen L. Williams et al., October, 1959, 53-61.
 Stinson, Pairlee J. et al., Sex Differences Among High-School Seniors, November, 1959, 103-108.
 Tamkin, Arthur S., A Survey of Educational Disability in Emotionally Disturbed Children, April, 1960, 313-315.
 Wall, Bartholomew D., Some Attitudinal Differences Among Educational Specialists, Administrators and Teachers, November, 1959, 115-117.
 Washburne, Norman F., Socioeconomic Status, Urbanism and Academic Performance on College, December, 1959, 130-137.
 What Teachers and Prospective Teachers Know About Phonics Generalizations, I. E. Aaron, May, 1960, 323-330.
 Where are They Now?--A Study of Teacher Supply in Mathematics and Science, Kenneth E. White, May, 1960, 331-335.
 White, Kenneth E., Where are They Now?--A Study of Teacher Supply in Mathematics and Science, May, 1960, 331-335.
 Williams, Helen L. et al., Stimulating Interest in Public Health Problems Among High School Pupils, October, 1959, 53-61.
 Wilson, J. A. R., Some Results of an Enrichment Program for Gifted Ninth Graders, December, 1959, 157-160.
 Witherspoon, Paul, A Comparison of the Problems of Certain Anglo- and Latin-American Junior High School Students, April, 1960, 295-299.
 Zeigler, Martin L. et al., A Study of the Effectiveness of a Summer Remedial Course in English for College Freshmen, October, 1959, 76-78.

